DOCUMENT RESUME

ED 305 383 TM 012 913

AUTHOR	Bethscheider, Janine K.
TITLE	The Perceptual Abilities Project. Technical Report No. 1988-4.
INSTITUTION	Johnson O'Connor Research Foundation, Chicago, IL. Human Engineering Lab.
PUB DATE	Oct 88
NOTE	89p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Adolescents; Adults; Cognitive Processes; *Perception; Perception Tests; Perceptual Development; *Spatial Ability; *Test Reliability; Test Validity; *Work Sample Tests
IDENTIFIERS	Flexibility of Closure; Johnson O Connor Aptitude Tests; *Spatial Tests; *Test Batteries

ABSTRACT

An experimental test battery designed to measure several perceptual abilities was administered to 1,368 (51.8% male) paying clients of the Johnson O'Connor Research Foundation (JOCRF) in an effort to identify and measure three perceptual abilities: (1) flexibility of closure; (2) speed of closure; and (3) spatial scanning. Subjects, who ranged in age from 14 to 69 years, also took the standard JOCRF battery. The experimental battery consisted of eight worksamples--six designed by the Educational Testing Service, and two developed within the JOCRF. Internal structure analyses of the tests indicated that the following worksamples had moderately high reliabilities: (1) Maze Tracing; (2) Hidden Patterns; (3) Drawing; and (4) Concealed Words. The experimental tests considered jointly were found to measure three distinct factors: (1) one represented by Gestalt Completion, Snowy Pictures, and Concealed Words; (2) a second measured by Hidden Patterns, Hidden Figures, Drawing, and Tactics; and (3) a third represented by Maze Tracing. The perceptual ability factors studied were not measured by the JOCRF standard battery. The Concealed Words Test, the Hidden Patterns Test, and Maze Tracing Speed Test were found to be the most suitable perceptual abilities measures; they may be useful additions to the standard JOCRF battery. Twenty-two tables and 16 figures present study data. (SLD)

ED305383

THE PERCEPTUAL ABILITIES PROJECT

Janine K. Bethscheider

JOHNSON O'CONNOR RESEARCH FOUNDATION

HUMAN ENGINEERING LABORATORY

Technical Report 1988-4

October 1988

TM012913

The Perceptual Abilities P·oject

Janine K. Bethscheider

ABSTRACT

An experimental test battery designed to measure several perceptual abilities was administered to over 1,300 examinees of the Johnson O'Connor Research Foundation (JOCRF). The battery consisted of eight worksamples: (a) Hidden Patterns, Hidden Figures, and Drawing, which were regarded as indices of flexibility of closure; (b) Gestalt Completion, Snowy Pictures, and Concealed Words, which were considered speed-of-closure tests; and (c) Maze Tracing Speed Test and Tactics, which were thought to be measures of spatial scanning. Six of the experimental tests were contained in the 1976 Kit designed by Educational Testing Service, and two worksamples were developed within the Foundation.

Internal-structure analyses of the individual perceptual abilities tests indicated that four of the worksamples (Maze Tracing, Hidden Patterns, Drawing, and Concealed Words) had moderately high reliabilities. The experimental tests considered jointly were found to measure three distinct factors: one represented by Gestalt Completion, Snowy Pictures, and Concealed Words; a second by Hidden Patterns, Hidden Figures, Drawing, and Tactics; and a third by Maze Tracing.

The relationships of the perceptual abilities to the aptitudes measured by the JOCRF standard battery were also investigated. The Maze Tracing Speed Test was not related to any of the Foundation worksamples. The speed-of-closure tests were moderately independent of the JOCRF worksamples. There was a relatively strong relationship between flexibility of closure and the spatially related worksamples in the standard battery, but the flexibility-of-closure measures also had a moderate amount of variance unique to themselves.

It was concluded that the constructs of flexibility of closure, speed of closure, and spatial scanning could not be measured sufficiently by the JOCRF standard battery alone. Moreover, on the basis of their psychometric properties, the Concealed Words Test, Hidden Patterns Test, and Maze Tracing Speed Test could be considered the most suitable of the perceptual abilities measures studied.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

This report describes a large-scale study conducted by the Johnson O'Connor Research Foundation (JOCRF) in an effort to identify and measure three perceptual abilities: flexibility of closure, speed of closure, and spatial scanning. The project utilized six tests that were designed by Educational Testing Service (ETS) to measure these three factors, along with two worksamples that were developed within the Foundation.

This report consists of four sections. This first section presents definitions of the three perceptual abilities under study and summarizes some of the literature in this area. Much of the material in this section was drawn from an unpublished report by Kaplan (1986). The second part of this report describes the eight worksamples that constitute the experimental perceptual abilities battery and discusses the types of analyses that were conducted for the study. The third section details the results of the analyses. This includes an assessment of the structure of the experimental worksamples, individually and as a complete battery, as well as an examination of the relationship of the perceptual abilities to the aptitudes measured by the JOCRF standard battery. These results are then summarized in the fourth section. Conclusions based on the findings of this study are also presented in the final part of this report.

The first person formally to study the domain of perceptual abilities was Thurstone (1944), who administered 40 tests representing a wide variety of perceptual processes to 194 university students. He found the tests to be described best by seven perceptual factors, including a speed-of-closure factor and a flexibility-of-closure factor (Tyler, 1956, pp. 228-230).

By the end of the 1940s, research activity related to the study of cognitive factors was rapidly increasing, especially with the advancement of factor-analytic techniques (Spearman, 1927; Thurstone, 1947). But the lack of any systematic procedures for identifying and defining aptitude factors made it difficult to make comparisons across studies. Consequently, it became virtually impossible to integrate all the findings into a comprehensive whole.

Therefore, in the early 1950s, ETS assembled a group of prominent researchers in the field to establish an integrated system for isolating and conceptualizing distinct cognitive factors. The outcome was the identification of 15 factors, of which flexibility of closure and speed of closure were two. Development of aptitude tests for each of the factors followed. Then, for each identified factor, those tests considered to be the best available measures of the factor were designated as marker, or reference, tests for that factor (i.e., tests believed to be measures of that one factor and only that factor). The result was the publication, in 1954, of the first edition of the Kit of Selected Tests for Reference Aptitude and Achievement Factors, which included three marker tests for each of the 15 factors. The purpose of the Kit was to "provide researchers with sets of common tests" (Ekstrom, French, Harman, & Dermen, 1976, p. 1), which would (a) facilitate comparisons across studies and (b) help clarify the contribution of other hypothesized measures of a factor.

In the late 1950s, a second conference was convened to consider the results of continued research on aptitude factors. This time, 24 factors were isolated, and a revised Kit of marker tests for the factors was published (French, Ekstrom, & Price, 1963). This second edition of the Kit included tests for spatial scanning, as well as flexibility of closure and speed of closure.

By the end of the 1960s, it appeared that a review of the status of the cognitive factors included in the 1963 Kit might be in order. Further updating and improvement of the Kit led to the publication of a third edition (Ekstrom et al., 1976). This latest Kit contained reference tests for 23 factors, again including flexibility of closure, speed of closure, and spatial scanning.

Definitions of the three perceptual abilities under study can be found in the following section. In addition, the literature on the relationships among the three factors, as well as their relationships to other cognitive abilities, is summarized.

Flexibility of Closure

In the most recent Manual for Kit of Factor-Referenced Cognitive Tests, flexibility of closure was defined as "the ability to hold a given visual percept or configuration in mind so as to disembed it from other well defined perceptual material" (Ekstrom et al., 1976, p. 19). In tests of this aptitude factor, the examinee must disembed a given figure from a more complex, and therefore distracting, perceptual field.

There are three marker tests for flexibility of closure in the 1976 Kit: the Hidden Patterns Test, the Hidden Figures Test, and the Copying Test. The two better-established measures, Hidden Patterns and Copying, not only present the examinee with the exact configuration to be found but also require the examinee to search for only one stimulus figure per item. Hidden Figures, on the other hand, requires the examinee to determine which one of five figures is embedded in each item. Because Hidden Figures has proven to be less clearly related to the other two markers than is desirable (Ekstrom, French, & Harman, 1979), ETS is in the process of revising this test so that the examinee must search for only one configuration in each item (Bejar & Yocum, 1986).

Speed of Closure

The definition posited for speed of closure in the 1976 Kit manual is: "the ability to unite an apparently disparate perceptual field into a single concept" (Ekstrom et al., 1976, p. 25). In tests of this factor, the examinee must take an ambiguous visual stimulus (e.g., one that is partially obliterated or "snow"-filled) and interpret it as a unitary whole.

There is some question about what differentiates measures of speed of closure from measures of flexibility of closure, but a major consideration appears to be the examinee's knowledge (or lack of knowledge) about the configuration for which he is searching. Typically, in speed-of-closure tests the examinee must close the disparate elements into a single unified figure without prior knowledge of the required configuration; in flexibility-of-closure tests the examinee is presented with the exact configuration to be found but must disembed it from a distracting perceptual field.

Two reference tests were included in the 1963 Kit for this factor: the Gestalt Completion Test and the Concealed Words Test. A third marker, Snowy Pictures, was added in the 1976 Kit. There is some evidence that these three measures are not reliable markers for the same factor (Ekstrom et al., 1979). One hypothesis is that there may be both semantic and perceptual speed-of-closure factors (Adcock & Martin, 1971; Messick & French, 1975), with Gestalt Completion and Snowy Pictures loading on the perceptual speed-of-closure factor and Concealed Words loading on the semantic or, possibly, both speed-of-closure factors (Ekstrom et al., 1979).

Spatial Scanning

This aptitude factor was defined in the 1976 Kit manual as "speed in exploring visually a wide or complicated spatial field" (Ekstrom et al., 1976, p. 155). The tests selected as markers for spatial scanning in the 1976 Kit were Maze Tracing Speed Test, Choosing a Path, and Map Planning Test. At least two of these reference tests have been used in a number of factor-analytic studies, but a clear spatial-scanning factor has emerged only twice (Ekstrom et al., 1979). In view of these findings, it has been concluded by ETS that revision of the marker tests for this factor may be necessary.

Relationships among Perceptual Abilities

A number of studies have investigated the relationship between flexibility of closure and speed of closure (Bechtoldt, 1947; Botzum, 1951; Glass, 1967; Messick & French, 1975; Mos, Wardell, & Royce, 1974; Pemberton, 1952; Thurstone, 1944). The majority testify to the existence of speed of closure as a factor distinct from flexibility of closure. For example, in a factor analysis conducted by Botzum (1951) on 46 cognitive tests, nine factors emerged, including two flexibility-of-closure factors and one speed-of-closure factor. Each of the flexibility-of-closure factors showed a minimal to moderately low correlation with the speed-of-closure factor (*r*s = .05 and .26, respectively). A factor analysis on 26 tests by Mos et al. (1974) yielded eight factors, including separate flexibility-of-closure and speed-of-closure factors. They also found a relatively low correlation (*r* = .21) between the two factors.

Several studies, however, have suggested some interdependence (i.e., correlation) between these two closure factors. Roff (1953), for instance, investigated the correlations between two measures of flexibility of closure (Hidden Patterns and Hidden Figures) and two measures of speed of closure (Gestalt Completion and Concealed Words). Hidden Figures correlated .35 with Gestalt Completion and .30 with Concealed Words; Hidden Patterns correlated .33 and .27 with Gestalt Completion and Concealed Words, respectively. Messick and French (1975) reported flexibility and speed of perceptual closure to be distinct albeit correlated factors (*r* = .29). Most of the evidence appears to indicate that although flexibility of closure is positively correlated with speed of closure, the two are separate factors.

Spatial scanning has been considered to be a factor quite distinct from flexibility of closure and speed of closure. Consequently, only one study specifically correlating spatial scanning with either closure factor was uncovered by the outside literature review. In this study, which was conducted by Guilford and Lacey (1947), Hidden Figures correlated .32 with Maze Tracing, .30 with Choosing a Path, and .28 with Map Planning. One study conducted by the Foundation also showed some correlation of one measure of spatial scanning (Map Planning) with one measure of flexibility of closure (Hidden Figures: *r* = .26; Technical Report 1981-2).

Relationships with Field Independence

In the 1940s, Witkin introduced the construct of field independence, measured by the Rod-and-Frame Test (RFT; Witkin, 1949). Researchers later discovered a substantial relationship between the RFT and embedded-figure types of tests (*r* = .37; Fry & Thompson, 1977; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962). This led to the acceptance of embedded-figure tests, as well as the RFT, as measures of field independence. More recently, however, after reviewing the research on field independence, Witkin and Goodenough (1981) concluded that the RFT represented field independence better than embedded-figure tests and that the two tests measured distinct aptitudes. This distinction was further supported by a correlational study conducted by Linn and Kyllonen (1981). Moreover, most of the recent evidence appears to indicate that although flexibility of closure and field independence are related, they are not identical factors. Currently, the majority of researchers regard as distinct the aptitudes measured by the RFT and embedded-figure tests, although some continue to use embedded-figure tests as a measure of field independence.

Studies relating field independence specifically to tests similar to Hidden Patterns or Copying were not uncovered in the outside literature review. Nor was research on the relationship of speed of closure to field independence. The independence of spatial scanning from field independence has never been questioned.

Relationships with Cognition and Personality

Flexibility of closure has been shown to correlate with a number of general cognitive aptitudes (Kastner, 1978; Messick & French, 1975). Parasuraman (1976) found that flexibility of closure was related to performance on tests of perceptual speed similar to the JOCRF Number Checking worksample. Barrett and Barker (1973) found a measure of flexibility of closure (Hidden Patterns) to be positively correlated to rate of reading. In a study conducted by the Foundation (Technical Report 1981-2), significant correlations emerged between flexibility of closure (as measured by Hidden Figures) and a number of other tests in either the JOCRF standard battery or the ETS *Kit*. Those findings can be summarized as follows. Correlations between Hidden Figures and measures of spatial ability ranged from .25 to .41 (*r*s = .41 with Surface Development, .35 with Paper Folding, .32 with Form Board, .26 with Wiggly Block, and .25 with Cube Comparisons). With tests of logical reasoning, Hidden Figures was found to correlate .29 with Diagramming Relationships and .28 with Analytical Reasoning and Inference.

Speed of closure has also been found to relate to a number of abilities in the cognitive domain. In particular, rate of reading, as well as comprehension, has been found to be positively correlated with speed of closure (Barrett & Barker, 1973; Baten, 1981; Glass, 1967).

An ETS study conducted by Ekstrom et al. (1979) utilized a battery of 33 tests that included three marker tests for each of 11 factors, including speed of closure. The correlations between the speed-of-closure measures were only low to moderate: Gestalt Completion was found to correlate .32 with Snowy Pictures and .26 with Concealed Words; Snowy Pictures correlated .19 with Concealed Words. In addition, distinct correlational patterns emerged for each of the speed-of-closure markers. Gestalt Completion showed moderate correlations only with the reference tests for spatial orientation (*r*s ranged from .24 to .36) and logical reasoning (*r*s ranged from .23 to .27). For Snowy Pictures, moderate correlations were found only with the indices of integrative processes (*r*s ranged from .19 to .39) and visual memory (*r*s ranged from .32 to .38). Concealed Words showed moderate correlational patterns with the markers of three factors: Verbal Closure (*r*s ranged from .17 to .41), Integrative

Processes (rs ranged from .16 to .32), and Logical Reasoning (rs ranged from .18 to .30). In a factor analysis, the three speed-of-closure tests had their highest loadings on the same factor; but Concealed Words also loaded on the Verbal Closure factor, and Snowy Pictures loaded on two other factors (Numerical Facility and Integrative Processes). These findings suggest that the three markers of speed of closure may be measuring somewhat different perceptual processes.

Spatial scanning has generally been considered an independent cognitive factor (Bunderson, 1967; Lemke, Klausmeier, & Harris, 1967). For instance, in a factor analysis of 34 task and ability tests (Lemke et al., 1967), seven ability factors emerged, including spatial scanning. In addition, spatial scanning was not substantially correlated with the concept-attainment and information-processing variables.

There is, however, some evidence that spatial scanning is moderately related to both a second-order visualization factor and a planning-ability factor (Royce, 1973). In addition, the results of a study conducted by Kastner (1978) indicated that spatial scanning was a contributor to concentration. Irons (1982) found the debugging skills of computer programmers to be related to performance on spatial-scanning measures. Finally, in the previously mentioned Foundation study (Technical Report 1981-2), significant correlations emerged between spatial scanning (as measured by Map Planning) and a number of other tests in the JOCRF battery or the ETS Kit. Those findings can be summarized as follows. Map Planning was found to moderately correlate with measures of spatial ability (rs = .43 with Surface Development, .38 with Form Board, .35 with Paper Folding and Cube Comparisons, and .31 with Wiggly Block). Correlations between Map Planning and indices of logical reasoning ranged from .30 to .36 (rs = .36 with Diagramming Relationships, .30 with Inference and Analytical Reasoning).

Few studies have investigated the relationships between perceptual processes and personality characteristics, although a number of typologies have been advanced over the years. One perception-personality typology that is frequently propounded contrasts analyzers with synthesizers. According to Tyler (1956), "the analytic observer concentrates on details and tends to see separate parts. The synthetic observer sees the field as an integrated whole but may miss some of its details completely" (p. 225). There is some evidence that flexibility of closure is related to analytic abilities, while speed of closure is related to synthetic (or holistic) abilities (Bouma & Ippel, 1983; Sawa, 1969).

Sex and Age Differences

Most of the literature on sex differences in perceptual abilities has focused on embedded-figure types of tests, such as Hidden Figures. For example, in their review of 32 studies that utilized embedded-figure tests, Maccoby and Jacklin (1974) found five that reported males significantly outperforming females and three that reported females significantly outperforming males. More recently, Shute, Pellegrino, Hubert, and Reynolds (1983) found that low-androgen males and high-androgen females (i.e., persons with hormone levels approaching those of the opposite sex) performed better on measures of flexibility of closure than high-androgen males and low-androgen females. Richmond (1980) reported no sex differences in scores on Hidden Figures or Copying. Likewise, in a Foundation study conducted by Trafton and Garrison (Techn cal Report 1982-3), no sex differences emerged for Hidden igures. Finally, in a meta-analysis of visuo-spatial tests including Hidden Figures, Linn and Petersen (1985) concluded that the reported sex effects for Hidden Figures and other tests in the same category were small, with most of the studies yielding no significant differences between males and females.

With regard to speed of closure and spatial scanning, research on sex differences was not uncovered in the outside literature review. In the Foundation study cited above, though, males were found to outperform females on Map Planning, a spatial-scanning measure.

Most of the research on age differences in perceptual abilities has addressed interactive rather than main effects. There is, however, some evidence of a gradual diminution of perceptual abilities with age, particularly for flexibility of closure (Crosson, 1984; Cunningham, 1978, 1981; Ohnmacht & Fleming, 1972; Ohnmacht, Weaver, & Kohler, 1970).

Other Conceptualizations

In the preceding sections, the ETS conceptualization of several perceptual abilities was presented. In this arrangement, Hidden Patterns, Hidden Figures, and Copying are considered indices of a factor labeled flexibility of closure; Gestalt Completion, Snowy Pictures, and Concealed Words are conceptualized as measures of a factor named speed of closure; and Maze Tracing Speed, Choosing a Path, and Map Planning are designated as markers for a spatial-scanning factor.

The foregoing conceptualization is the most common one used by social scientists, but that does not preclude other arrangements of these nine ETS reference tests. For instance, W. Wothke (personal communication, December 16, 1985) has suggested the following four-factor ordering:
(a) embedding-disembedding, as measured by Hidden Patterns and Hidden Figures; (b) closure over random noise, as measured by Gestalt Completion, Snowy Pictures, and Concealed Words;
(c) route planning/map reading, as measured by Maze Tracing Speed, Choosing a Path, and Map Planning; and (d) structural visualization, as partially measured by Copying.

Review of the extant literature did not uncover any studies that investigated the relationships among all three of flexibility of closure, speed of closure, and spatial scanning. Without the support of empirical evidence, neither the distinctness of the three perceptual abilities nor ETS's conceptualization of the nine aforementioned markers as measures of flexibility of closure, speed of closure, or spatial scanning should be regarded as definitive.

METHOD

Examinees

The examinees for this study were paying clients of the JOCRF. These examinees came to the Foundation for testing in order to obtain information about their aptitudes that they could use in making educational and occupational decisions. The JOCRF examinee population is a relatively homogeneous group with respect to education and socioeconomic status. Foundation clients typically are white and middle- to upper-middle-class. The great majority are college-bound or college-educated.

A total of 1,368 examinees completed one or more of the worksamples in the experimental battery. Of these, 708 were male (51.8%) and 660 were female (48.2%). The age of the clients ranged from 14 to 69, with an average age of 24.9 (SD = 9.6) and a median age of 20. Approximately 54% of the examinees were tested in Foundation offices in the southern United States (Houston, Dallas, New Orleans), 27% in test centers in the East (Philadelphia, Washington, DC), and 19% in offices located in the West and Northwest (Denver, Seattle).

Measures

The JOCRF Standard Aptitude Battery

The perceptual abilities battery was administered in conjunction with a larger aptitude test battery offered by the Foundation. A brief description of the aptitudes measured by the JOCRF standard worksamples is contained in Table 1. A more detailed description of the tests in the experimental perceptual abilities battery follows.

The Experimental Perceptual Abilities Battery

The experimental battery consisted of eight worksamples. three thought to be measures of flexibility of closure, three of speed of closure, and two of spatial scanning. Six of the tests were contained in the 1976 *Kit* designed by ETS, and two worksamples were developed within the Foundation. In this and subsequent sections of the report that consider the experimental worksamples individually, first the speed of closure markers will

Table 1

Aptitudes Measured in the JOCRF Standard Battery

Name	Description
Graphoria	Speed and accuracy in noticing if pairs of numbers are the same or different.
Ideaphoria	Measure of verbal fluency, the rate of flow of ideas.
Foresight	Ability to keep one's mind on a long-range goal.
Inductive Reasoning	Quickness in seeing a common element among separate facts, ideas, or observations.
Analytical Reasoning	Quickness in arranging ideas into a logical sequence.
Wiggly Block	Measure of structural visualization, an aptitude for visualizing three-dimensional forms. Measured by the ability to reconstruct a three-dimensional block.
Paper Folding	A second measure of structural visualization. Measured by the ability to rotate two-dimensional surfaces through three-dimensional space.
Personality	Tendency to react from a general, objective viewpoint versus reacting from a personal, subjective viewpoint. Describes how well-suited a person is for work that is highly oriented toward person contact (Objective) or toward individual performance (Subjective).
Tonal Memory	Ability to remember sequences of tones.
Pitch Discrimination	Ability to differentiate fine differences in pitch.
Rhythm Memory	Ability to remember complex rhythmic patterns.
Memory for Design	Memory for straight-line patterns.
Silograms	Associative memory for English words paired with nonsense syllables.
Number Memory	Ability to remember several six-digit numbers simultaneously.
Observation	Quickness in recalling fine visual details.
Finger Dexterity	Speed and accuracy in manipulating small objects with one's fingers.
Tweezer Dexterity	Speed and accuracy in handling small objects with tweezers.

be discussed, then the indices of flexibility of closure, and, last, the measures of spatial scanning.

Gestalt Completion Test (Worksample 729 A*). This test is an ETS marker of speed of closure, with reported reliabilities ranging from .77 to .85 (Ekstrom et al., 1976). The worksample consists of two parts, each lasting two minutes and containing 10 items. In the Gestalt Completion Test, the examinee is shown drawings composed of black blotches that represent parts of objects and is asked to identify these incompletely drawn figures. Figure 1 contains a sample item from this test. (The examples displayed in this report for all the tests except Drawing are not items from the tests themselves but items that are presented in conjunction with the instructions for the test for purposes of practice or illustration.)

Figure 1

Sample Item from the Gestalt Completion Test

From Manual for Kit of Factor-Referenced Cognitive Tests (p. 27) by R. B. Ekstrom, J. W. French, H. H. Harman, and D. Dermen, 1976, Princeton, NJ: Educational Testing Service. Copyright © 1975 by Educational Testing Service. Reprinted by permission.

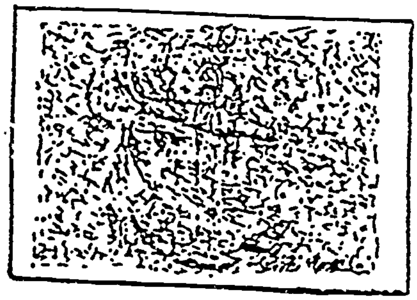
The answer is: flag.

Snowy Pictures (Worksample 732 A*). This test also is an ETS reference test for the speed-of-closure factor, with a reported reliability of .68 (Ekstrom et al., 1976). It comprises two parts, each lasting four minutes and containing 12 items. In Snowy Pictures, the examinee is presented with pictures of objects covered with splatters and asked to identify these partially obliterated figures. An example of an item from this test is shown in Figure 2.

Figure 2

Sample Item from Snowy Pictures

From Manual for Kit of Factor-Referenced Cognitive Tests (p. 29) by R. B. Ekstrom, J. W. French, H. H. Harman, and D. Dermen, 1976, Princeton, NJ: Educational Testing Service. Copyright © 1975 by Educational Testing Service. Reprinted by permission.



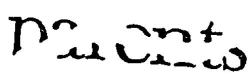
The answer is: anchor.

Concealed Words Test (Worksample 728 A*). Likewise an ETS marker of speed of closure, this test consists of two parts, each having 25 items and lasting four minutes. The reported reliabilities for this worksample range from .63 to .83 (Ekstrom et al., 1976). In the Concealed Words Test, the examinee is shown words containing partly obliterated letters and is asked to identify the complete words. Figure 3 contains an example of an item from this test.

Figure 3

Sample Item from the Concealed Words Test

From Manual for Kit of Factor-Referenced Cognitive Tests (p. 28) by R. B. Ekstrom, J. W. French, H. H. Harman, and D. Dermen, 1976, Princeton, NJ: Educational Testing Service. Copyright © 1962, 1975 by Educational Testing Service. Reprinted by permission.



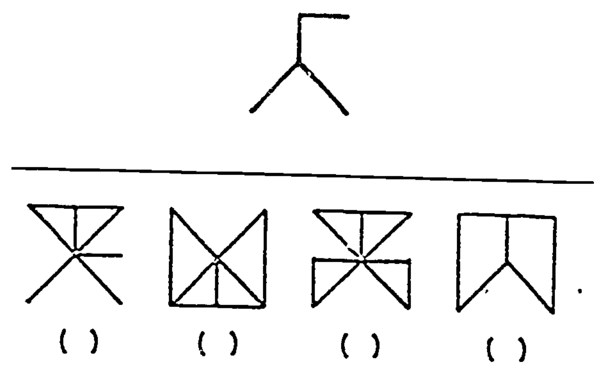
The answer is: parents.

Hidden Patterns Test (Worksample 727 A*). This test is an ETS marker of flexibility of closure, with reported reliabilities ranging from .81 to .91 (Ekstrom et al., 1976). The worksample contains two parts, each comprising 200 items and lasting three minutes. In the Hidden Patterns Test, the examinee is given one simple geometric figure (or model) and is asked to determine whether that figure is embedded in each of a series of slightly more-complex geometric patterns. The items on this worksample are relatively easy but must be solved under highly speeded conditions. Figure 4 displays a model figure along with several examples of items from this test.

Figure 4

Model Figure and Sample Items from the Hidden Patterns Test

From Manual for Kit of Factor-Referenced Cognitive Tests (p. 22) by R. B. Ekstrom, J. W. French, H. H. Harman, and D. Dermen, 1976, Princeton, NJ: Educational Testing Service. Copyright © 1962, 1975 by Educational Testing Service. Reprinted by permission.



The answers are: the model appears in the first, third, and fourth patterns.

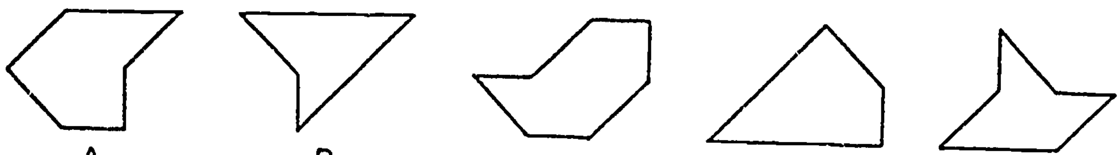
Hidden Figures Test (Worksample 730 A*). Also an ETS marker for the flexibility of closure factor, this test has reliabilities reported in the .80 to .83 range (Ekstrom et al., 1976). Hidden Figures has two parts, each with 16 items and lasting 12 minutes. Due to time constraints, however, only the first half of this test was administered as part of the perceptual abilities battery. In the Hidden Figures Test, each item consists of a complex geometrical pattern. The examinee is given five figures and must decide which of the five is embedded in each of the patterns. The items on this test are very difficult. Figure 5 presents an example of five figures and two items from this test.

Drawing (Worksample 494 B*). This worksample is a Foundation test similar to the ETS Copying Test, which is a marker of flexibility of closure. Both the JOCRF Drawing test and the ETS Copying test are adaptations of the Copying subtest in the MacQuarrie Test for Mechanical Ability, which was developed in 1925. In the MacQuarrie Copying subtest, the examinee is presented with drawings consisting of four straight lines and asked to copy each pattern onto a five-by-five grid of dots. One dot on each matrix is circled to denote the starting point for the drawing. Likewise, in the ETS Copying test, the examinee is asked to copy four-line drawings onto five-by-five grids of dots.

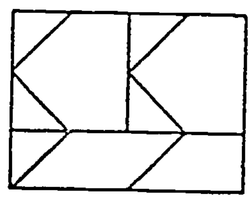
Figure 5

Sample Figures and Items from the Hidden Figures Test

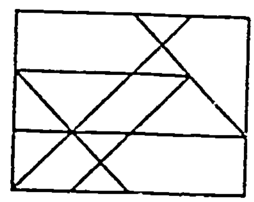
From Manual for Kit of Factor-Referenced Cognitive Tests (p. 21) by R. B. Ekstrom, J. W. French, H. H. Harman, and D. Dermen, 1976, Princeton, NJ: Educational Testing Service.



A B C D E



A B C D E



A B C D E

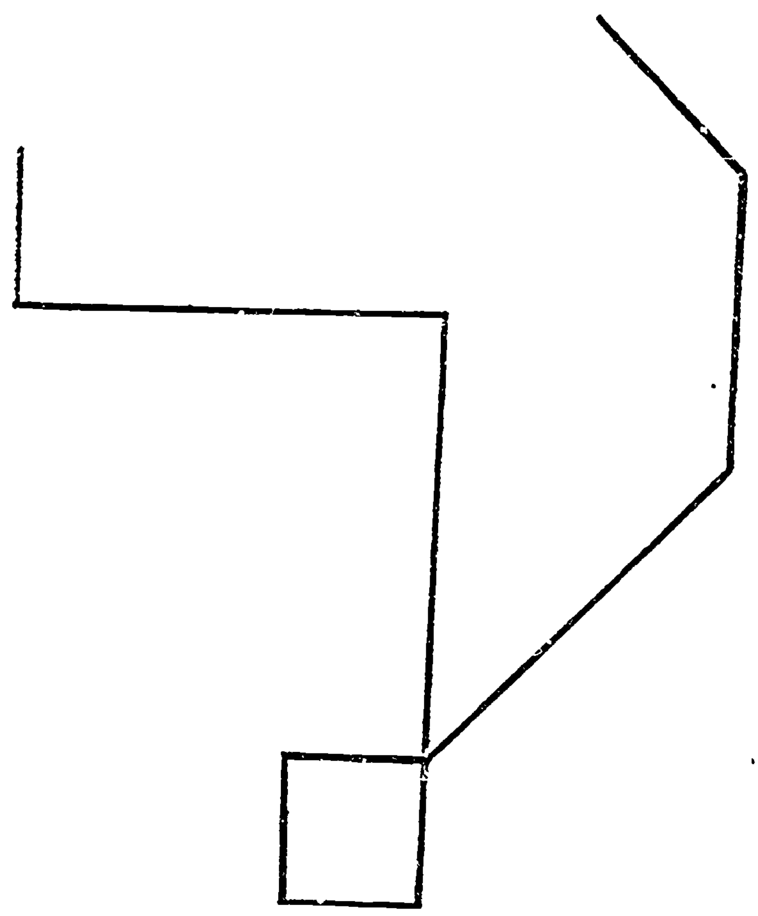
The answers are: A and D, respectively.

The original Foundation version (Form A), which was developed in the late 1940s, utilized eight drawings, each composed of 12 lines (Hill, 1970). This worksample proved to be too difficult for examinees, however, and a second version with eight eight-line drawings (Form B) was subsequently designed (Hill, 1970). The current version of Drawing (Form B*) comprises 10 items (two newly designed drawings for the first two items plus the eight from Form B), lasting one minute apiece. Each item consists of a drawing of eight or nine straight lines. While the drawing is projected on a screen, the examinee is asked to copy the pattern onto a sheet of paper displaying a nine-by-nine grid of dots. One line of the drawing is already drawn on the answer sheet to serve as a reference line. The solution (i.e., the drawing superimposed on the matrix of dots) is shown after each of the first two items. A sample item from the Drawing test is displayed in Figure 6.

The Drawing worksample was designed by the Foundation as a paper-and-pencil test of structural visualization (Minutes of Meeting of the Trustees, 1964). Correlations between Drawing and other structural visualization measures in the JOCRF aptitude battery ranged from .44 to .47 (rs = .47 with Black Cube, .44 to .46 with Wiggly Block; Minutes of Meeting of the Trustees, 1966;

Figure 6

Sample Item from Drawing



Minutes of Meeting of the Research Department and Executive Committee, 1971). Reliabilities of .86 (Wallace & Mandell, 1972) and .89 (Poff & Holder, 1974) have been reported for the Drawing worksample.

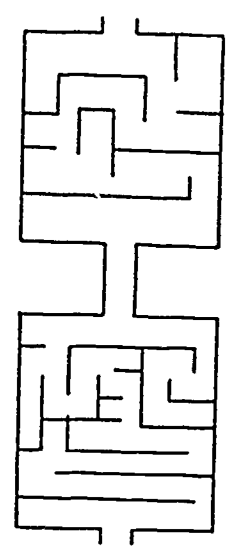
Two distinctions between copying-type tests, such as those by MacQuarrie and ETS, and the Foundation Drawing test are worth mentioning. First, in the copying tests the pattern to be reproduced is presented on paper immediately to the left of the matrix of dots, whereas in the Drawing test the pattern is projected on a screen and only the grid of dots appears on paper. Second, in the Drawing test the examinee must reproduce to scale the drawings appearing on the screen, whereas in the copying-type tests each line drawn by the examinee must be the same length as the corresponding line of the pattern to the left of the grid (i.e., scaling is not required).

<u>Maze Tracing Speed Test</u> (Worksample 731 A*). This test is an ETS marker of spatial scanning, with reported reliabilities of .89 to .94 (Ekstrom et al., 1976). The worksample consists of two parts, each lasting three minutes. In each part, the examinee is asked to trace a path through a series of 24 mazes. Examples of items used in the test are shown in Figure 7.

Figure 7

<u>Sample Items from the Maze Tracing Speed Test</u>

From <u>Manual for Kit of Factor-Referenced Cognitive Tests</u> (p. 157) by R. B. Ekstrom, J. W. French, H. H. Harman, and D. Dermen, 1976, Princeton, NJ: Educational Testing Service. Copyright © 1962 by Educational Testing Service. Reprinted by permission.

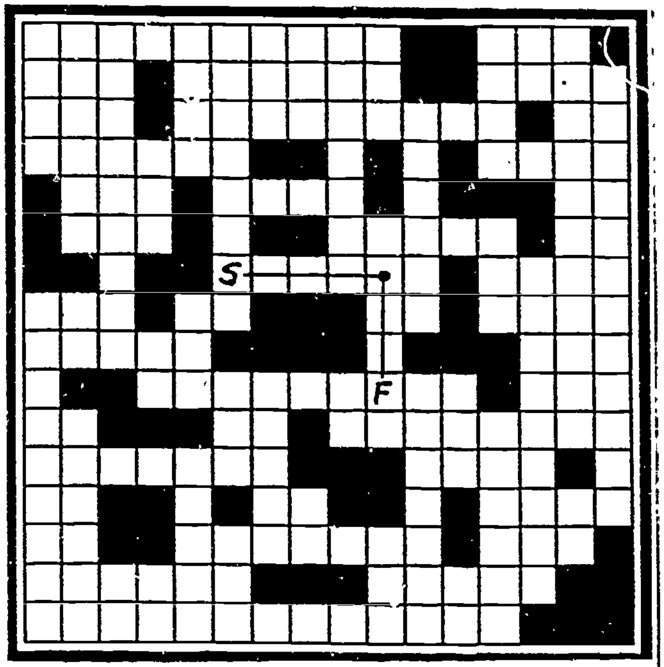


<u>Tactics</u> (Worksamples 413 A* and 413 B*). This worksample is a Foundation test similar to the ETS Map Planning Test, which is a marker of spatial scanning. In each item, the examinee is presented with a 16-by-16 partially blockaded checkered board and

is asked to find, as quickly as possible, the most efficient route from one designated square to another. Figure 8 contains an example of a Tactics item.

Figure 8

Sample Item from Tactics



Tactics has been studied less than Drawing. The original version of Tactics (Form A) was developed in the mid-1940s and consisted of five trials (Technical Report 168). In the late 1960s, the five items were revised or replaced, with the new version labeled Form BA (Johnson O'Connor Research Foundation, 1968). For the current study, two forms of Tactics were utilized: a five-trial version (Wks. 413 A*), consisting of the items from Form BA, and a 10-trial version (Wks. 413 B*), composed of the five items from Form A* plus five newly created items.

Analyses

The objectives of the data analyses for this study were (a) to evaluate the psychometric properties of the individual worksamples in the experimental perceptual abilities battery, (b) to investigate sex and age differences in performance on the individual worksamples, (c) to assess the relationships among the worksamples within the experimental battery, (d) to examine the relationships of the perceptual abilities to the aptitudes measured by the JOCRF standard battery, and (e) to explore the relationships of the perceptual abilities to laterality, education, and college major. To address these objectives, five series of analyses were performed on the data collected in this study. Unless otherwise stated, the SPSS/PC+ (version 2.0; Norusis, 1988a) and SPSS/PC+ Advanced Statistics (version 2.0; Norusis, 1988b) computer software packages were used for these analyses.

Item Analyses

The initial series of analyses examined the internal structure of each test in the experimental battery. For this study, the item analyses generally included calculation of the following statistics: (a) reliability coefficient, (b) item difficulties, (c) item-total correlations, (d) Rasch fit statistics, and (e) item factor analysis. A review of the basic statistical techniques relevant to these item analyses is contained in the following paragraphs.

Analysis of the internal structure of cognitive tests typically includes an internal consistency estimate of reliability. Coefficient alpha is one of the most commonly used indices of internal consistency for tests that are not appreciably speeded. A formula for deriving an estimate of alpha reliability is shown below:

$$r_{kk} = \left(\frac{k}{k-1}\right)\left(\frac{\sigma_y^2 - \Sigma\sigma_i^2}{\sigma_y^2}\right),$$

where r_{kk} is coefficient alpha, k is the number of items in the worksample, σ_y^2 is the variance of the total score, and σ_i^2 is the variance of an item score (Nunnally, 1967, p. 196).

With speeded tests, where most of the items attempted are answered correctly, a reliability estimate based on item-level data is not appropriate. An alternative is to calculate the reliability using subtest scores, provided the division of the worksample into subtests is made on the basis of time rather than items. In this study, for the experimental tests composed of two separately timed parts, the reliability was computed as the split-half correlation between the two parts of the test, which was then adjusted by the Spearman-Brown correction, yielding the reliability estimate for the complete worksample. The Spearman-Brown formula for split-half reliability is as follows:

$$r_{kk} = \frac{2\, r_{12}}{1 + r_{12}},$$

where r_{kk} is the reliability of the whole worksample and r_{12} is the correlation between Parts 1 and 2 of the test (Nunnally, 1967, p. 194).

An acceptable reliability estimate generally falls in the .80s or .90s (Anastasi, 1976; Test Information Bulletin 1980-7). A close-to-adequate reliability (.70 or above) indicates that with further refinement the worksample could be acceptable. A reliability coefficient below .70 suggests that the test needs substantial improvement before it could be incorporated into the JOCRF standard battery.

One way to increase a test's reliability is to lengthen it by adding items. The number of items necessary to obtain a specified reliability can be estimated using an inversion of the Spearman-Brown formula (Nunnally, 1967). For this study, the estimated lengthening required to obtain a reliability coefficient of .80 was calculated for any worksample in the experimental battery that had a lower-than-desirable reliability.

The second step in item analysis for this study was to determine item difficulties. The difficulty of an item (p value) is defined as the proportion of examinees who correctly completed the item. The larger the p value, the easier the item was for those tested. For example, a p value of .95 means that 95% of the examinees answered the item correctly.

It is important to measure item difficulties in order to discover whether the test's distribution of item difficulties is suitable for the distribution of ability levels of the examinees. As stated by Anastasi (1976):

> Ability tests are designed to assess as accurately as possible each individual's level of attainment in the particular ability. For this purpose, if no one passes an item, it is excess baggage in the test. The same is true of items that everyone passes. Neither of these types of items provides any information about individual differences. Since such items do not affect the variability of test scores, they contribute nothing to the reliability or validity of the test. The closer the difficulty of an item approaches 1.00 or 0, the less differential information about examinees it contributes. (p. 199)

If the distribution of item difficulties shows that an aptitude test has too few items with large p values (or too many items with small p values), it means that the test is too difficult and cannot properly differentiate examinees at the lower end of the ability range. If a test lacks items with small p values (or contains an abundance of items with large p values), it indicates that the test is probably too easy and cannot discriminate properly at the higher end of the ability range.

A Rasch plot of person-ability and item-difficulty distributions from the MSCALE program (version 2.0; Wright, Congdon, & Rossner, 1987) was also examined to assess the distribution of item difficulties across the range of person ability. It is usu y considered desirable to have a test with a wide distribution of item difficulties and an average difficulty of .50.

The third type of item statistic calculated in this study was the item-total correlation, i.e., the correlation of examinees' scores on a particular item with their total test scores. In calculating an item-total correlation, however, one must remember that the item is a part of the total test. If no adjustment is made for this part-whole overlap, the resulting item-total correlation is spuriously high. Correlating the item score with total score less the item score (i.e., the total of the other item scores) was the procedure used in this study to correct for the item-total overlap effect.

An item-total correlation "describes the degree to which an individual item fits the test; e.g., the degree to which the item measures the same latent trait that is measured by the other items on the test" (Technical Report 1987-1, p. 24). Items that correlate most highly with total score usually fit the test best and add the most to a worksample's reliability. As noted by Tal (Technical Report 1987-1):

> Good-fitting items increase a test's reliability, while poorly fitting items tend to decrease reliability or to add very little to reliability; i.e., removal of a good item decreases the reliability of a test, while removal of a bad item increases reliability or leaves it the same. (p. 24)

In general, item-total correlations above .30 are considered satisfactory.

Rasch fit statistics also were utilized in this study to assess the quality of a test's items. The Rasch model is a latent-trait model that produces item fit ("infit") statistics that can be used to determine which items are good indicators of ability. The infit statistic for an item represents the correspondence (or fit) between the set of examinees who answered the item correctly and the set of examinees who would be expected to answer the item correctly based on their ability estimates. In the Rasch model, large negative infit values mean very good item fit, infit values in the range of -2 to +2 indicate average fit, and large positive infit values imply poor item fit. The fit statistics for this study were estimated using the <u>MSCALE</u> program (version 2.0; Wright, Congdon, & Rossner, 1987).

The final item-analysis procedure used in this study was exploratory item factor analysis to verify that the set of test items could be adequately explained by a single factor. Estimates of the initial factors were obtained from principal components analysis. The decision regarding the number of factors needed to represent the item data was based on eigenvalues and the percentage of total variance accounted for by each extracted factor. A scree plot of the total variance associated with each factor was also examined.

Analyses of Sex and Age Effects

Sex and age differences in performance on the individual worksamples in the experimental battery were evaluated using the classical experimental approach to analysis of variance of the test scores by sex and age. Analysis of variance was used to answer three questions:

1. What is the main effect of sex on performance scores when differences in age are controlled?
2. What is the main effect of age on performance when scores are adjusted for differences in sex?
3. Is there an interactive effect of sex and age on scores?

For these analyses, the sample was partitioned into the following nine age groups: (a) 14-16, (b) 17-19, (c) 20-22, (d) 23-25, (e) 26-30, (f) 31-35, (g) 36-40, (h) 41-45, and (i) 46 and older. The resulting 2 x 9 designs contained unequal sample sizes in the cells.

The size of each significant main effect was assessed by estimating omega-squared ($\underline{w}^2$), which represents the proportion of variance in the worksample that was attributable to the given effect. The square root of omega-squared (i.e., $\sqrt{\underline{w}^2}$) is equivalent to the effect of a correlation of the same value.

Analyses of Perceptual Abilities Battery

In the third phase of analyses, the structure of the complete perceptual abilities battery was evaluated by investigating the pattern of correlations among the experimental worksamples. First, the Pearson product moment correlation coefficient $\underline{r}$ (zero order correlation) was used to measure the magnitude of the relationships among the eight perceptual abilities worksamples. Then, exploratory factor analyses of the correlation matrix were undertaken to discover the underlying factors that help explain the observed correlations among the worksamples within the experimental battery. For a factor model to be appropriate, however, the measures must be related to each other to a nontrivial extent. If correlations among the worksamples are small, it is unlikely that they share common factors.

Estimates of the number of factors necessary to represent the data were obtained from principal components analysis. Generally, only factors with an eigenvalue greater than unity were considered salient. The scree plot served as an additional criterion for ascertaining the significant factors.

After the number of salient factors was determined, those factors were extracted from the correlation matrix using principal components analysis. Because it is usually difficult to identify meaningful factors based on extraction alone, varimax rotation was then performed in order to enhance the interpretability of the extracted factors. After rotation, the matrix of factor loadings was examined.

The factor loadings represent the correlations between the factors and the worksamples. Typically, interpretation of each factor is based on those worksamples with large loadings on the factor (.40 or above). For the factor analysis to be considered successful, few, if any, of the worksamples should have large loadings on more than one factor. Furthermore, each factor should have large loadings for only some of the worksamples. The results should then be factors that are clearly differentiated and easily interpretable.

Analyses of Perceptual Abilities and JOCRF Batteries

In the fourth set of analyses for this study, the relationships of the perceptual abilities to the aptitudes measured by the Foundation's standard battery were explored. First, zero-order correlations between the perceptual abilities and the JOCRF worksamples were computed to determine their degree of relationship. Next, several exploratory factor analyses were performed. Estimates of the number of factors necessary to represent the data were obtained from principal components analysis. Extracted factors were then rotated to the varimax criterion and the resulting factor solutions assessed in terms of interpretability.

Multiple regression analyses were then performed on the data to determine whether the experimental tests showed considerable redundancy with worksamples in the standard battery or measured constructs unique to the perceptual abilities battery. The larger the multiple R, the more overlap or redundancy between the experimental test and the Foundation worksamples. The size of the effect was assessed by the square of the multiple R (i.e., R^2), which represents the amount of variance in the perceptual abilities test that is explained by the standard worksamples. If perceptual abilities scores can be well-predicted from scores on the Foundation worksamples, it means scores on the experimental tests can be obtained more efficiently--that is, without having to administer the perceptual abilities battery. Consequently, unless a perceptual abilities measure makes a unique contribution to the JOCRF battery, its inclusion in the standard battery would be of limited value.

Analyses of Perceptual Abilities and Laterality, Educational Level, and College Major

To assess the relationships of laterality, educational level, and college major to the perceptual abilities, three multivariate analyses of variance were performed. For each analysis, the dependent variables were the experimental measures with the exception of Tactics. (Including Tactics would have meant that those individuals who were administered the 5-trial version of Tactics would be excluded from the analysis, thereby considerably reducing the sample size.) For the Tactics test, univariate analyses of variance were performed. For each significant effect on a multivariate analysis of variance, a univariate analysis of variance was then performed to determine which of the perceptual abilities tests showed the effect.

In the first multivariate analysis of variance, the independent variables were eyedness and handedness along with sex, which has been found to interact significantly with laterality (Technical Report 1987-2). Eyedness scores were computed as the ratio of the number of trials on which the right eye was used as the sighting eye to the total number of trials administered on the eyedness portion of the Foundation's Eye & Hand worksample. Handedness scores were computed as the ratio of the number of trials on which the right hand was used to the total number of trials on the handedness section of the Eye & Hand worksample. Examinees were classified as (a) completely left-eyed if their eyedness ratio was 0, (b) variable-eyed if their eyedness ratio was between 0 and 1, or (c) completely right-eyed if their eyedness ratio was 1. (Individuals could be classified as variable-eyed because they spontaneously switched eyes in Parts 1 or 2 of the worksample or they switched eyes when directed to use their nondominant eye in Part 3.) The sample was also partitioned into three handedness groups: (a) primarily left-handed (for a handedness ratio less than .50), (b) variable-handed (for a ratio of .50 through .89), or (c) essentially right-handed (for a ratio of .90 to 1).

In the second multivariate analysis of variance, the independent variable was the educational level of examinees at the time of their testing. For the third multivariate analysis, examinees' major field of study was the independent variable. For this analysis, the following college-major categories were employed: (a) engineering, (b) physical sciences, (c) biological sciences, (d) business, (e) social sciences, (f) humanities, and (g) education. Only individuals who had completed at least two years of college were included in this analysis.

The results of the analyses for this study are reported in the following sections.

RESULTS AND DISCUSSION

The Experimental Worksamples Considered Individually

Gestalt Completion Test

Scores. As mentioned earlier, the Gestalt Completion Test was divided into two comparable parts. An examinee's score for a particular part was the number of items he or she correctly answered within that part. A score for the entire test was computed by adding together an examinee's part scores. Table 2 presents descriptive statistics for each part of Gestalt Completion.

Table 2

Descriptive Statistics for Gestalt Completion Test

Section	No. of items	Mean	*SD*	Range
Part 1	10	7.7	1.4	0-10
Part 2	10	6.6	2.0	0-10
Total	20	14.3	2.9	0-20

Note. *N* = 1,351.

Total scores on the worksample ranged from the test floor (i.e., 0) to the test ceiling (i.e., 20). Although there were no appreciable ceiling or floor effects, the distribution of test scores was asymmetrical, with scores concentrated toward the high end of the distribution.

Internal structure. Analysis of the internal structure of the Gestalt Completion Test was straightforward. The split-half correlation between Parts 1 and 2 of the worksample was .52. Applying the Spearman-Brown correction to the correlation resulted in a reliability estimate of .68, which is lower than is desirable by Foundation standards. To increase the reliability to an acceptable .80 would require almost doubling the length of the worksample (from 20 items to 38 items), which would increase total test time from 4 to 7.5 minutes.

Table 3 displays the item difficulties, item-total correlations, and infit statistics for the test. Item difficulties ranged from a low *p* value of .03 for Item 7 to a high of .98 for Items 3 and 4, with an average item difficulty of .71. It is often appropriate to have a few easy problems at the beginning of a test to serve as warm-up items, but too many simple problems is inefficient. In Part 1 of Gestalt Completion, the first four items (or 40% of that part) had *p* values in excess of .95, indicating that there probably are too many extremely easy items, which may contribute to the less-than-adequate split-half reliability of the worksample.

Item-total correlations ranged from .11 (Item 3) to .54 (Item 11). Items 1 through 4 and 7 showed low item-total correlations. In addition, Items 5, 12, 14, 15, 16, 18, and 20

Table 3

Item Statistics for Gestalt Completion Test

Item	Difficulty (p value)	Item-total correlation	Fit statistic
1	.97	.20	-1.03
2	.96	.16	-.29
3	.98	.11	-1.30
4	.98	.15	-1.29
5	.92	.21	.62
6	.86	.32	.27
7	.03	.15	-.90
8	.80	.37	-.29
9	.48	.39	-1.22
10	.67	.43	-2.12
11	.78	.54	-5.83
12	.86	.27	1.12
13	.57	.35	1.49
14	.31	.24	3.12
15	.86	.27	1.16
16	.94	.29	-.86
17	.66	.52	-6.29
18	.84	.24	2.37
19	.64	.40	-.82
20	.09	.24	-1.41

Note. N = 1,351.

displayed relatively low item-total correlations. Eight items correlated at least .30 with total score.

Infit statistics for the Gestalt Completion Test ranged from -6.29 to 3.12. Items 14 and 18 showed moderately poor fit relative to the expected pattern of responses. The other items displayed average or better fit.

Principal components factoring of the item correlation matrix extracted six factors with eigenvalues greater than or equal to 1.0, namely, 3.48, 1.34, 1.12, 1.09, 1.04, and 1.00. The first factor accounted for 17.4% of the variance in the test, while the second accounted for 6.7%. None of the remaining factors accounted for more than 6% of the variance. The scree plot indicated that a one-factor model was sufficient for the Gestalt Completion Test.

<u>Sex and age effects</u>. The analysis of variance of test scores by sex and age group yielded a significant main effect for sex for this worksample, $F(1;1,296) = 18.50$, $p < .001$. The main effect for age was also significant, $F(8;1,296) = 5.77$, $p < .001$. The interaction between sex and age was not significant.

Males did better than females on Gestalt Completion, averaging 14.68 correctly completed items to the females' average of 13.94. When scores are adjusted for age, the difference in performance between males and females is .23 of a standard deviation. The proportion of variance attributable to sex alone is small ($w^2 = .01$), which is equivalent to a correlation of .10.

When scores are adjusted for sex, the proportion of variance in Gestalt Completion that is accounted for by age alone is .03. This is comparable to the effect of a correlation of .17.

The age curve for this worksample is graphically depicted in Figure 9. The plotted values are the mean scores (in deviation units) of the nine age groups, adjusted for the effects of sex. No growth curve is manifested in Figure 9; that is, the adult plateau for Gestalt Completion is in evidence by age 14 and appears to extend into the mid-30s. A mild decrease and leveling off is observed for examinees in their late 30s and early 40s, with a sharper decline observed for those 46 and older.

<u>Summary</u>. The main findings with regard to the Gestalt Completion Test can be summarized as follows:

1. The split-half reliability of .68 for Gestalt Completion was lower than is desirable for a test reliability.

2. The skewed distribution of test scores and the relatively high average item difficulty of .71 suggest that the worksample probably contains too many easy items for the Foundation sample.

3. Forty percent of the items correlated at least .30 with total score.

4. None of the items showed extremely poor fit.

5. Significant sex and age effects were found for this worksample, with males scoring about one-fourth of a standard deviation higher than females.

Figure 9

Age Curve for Gestalt Completion Test

Note. Plotted values are not smoothed.
[a]Adjusted for sex.

Snowy Pictures

Scores. As noted earlier, Snowy Pictures was divided into two parallel parts. The number of items answered correctly within a part determined an examinee's score for that part. A score for the whole test was obtained by totaling an examinee's scores on Parts 1 and 2. Descriptive statistics for each part of Snowy Pictures are displayed in Table 4.

Table 4

Descriptive Statistics for Snowy Pictures

Section	No. of items	Mean	SD	Range
Part 1	12	9.8	1.8	2-12
Part 2	12	6.6	2.2	0-12
Total	24	16.4	3.4	4-24

Note. *N* = 1,350.

Total scores on the worksample ranged from 4 to 24, the maximum possible. Although there were no appreciable ceiling or floor effects, the distribution of test scores was asymmetrical, with scores concentrated toward the high end of the distribution.

Internal structure. For Snowy Pictures, the correlation between Parts 1 and 2 was .44, yielding a reliability coefficient of .61, which by Foundation standards is lower than desirable. Increasing the reliability to .80 would require lengthening the worksample 2.6 times, resulting in a 62-item test lasting 21 minutes.

The item difficulties, item-total correlations, and infit statistics for the test are presented in Table 5. The *p* values ranged from .15 to .99, with an average item difficulty of .68. Two items from Part 1 (1 and 11) were extremely easy for the Foundation sample. One easy item at the beginning of the test may be useful as a warm-up item, but a second easy item near the end of Part 1 serves no clear purpose.

Item-total correlations ranged from a low of .04 to a high of .39. Three items from Part 1 (1, 4, and 9) had extremely low item-total correlations; two items from Part 2 (21 and 22) displayed low item-total correlations. Seven items correlated at least .30 with total score. Infit statistics ranged from -4.56 to 5.87. Items 4 and 21 exhibited extremely poor fit, while Items 9 and 17 showed moderately poor fit relative to the expected pattern of responses. The other items displayed at least average fit.

As described earlier, each Snowy Pictures item consists of a figure embedded in random splatters or "snow," which the examinee must interpret as a unitary whole. Common to all three items that were misfitting due to their low correlations with total score (i.e., Items 4, 9, and 21) is an ambiguity regarding what

Table 5

Item Statistics for Snowy Pictures

Item	Difficulty (p value)	Item-total correlation	Fit statistic
1	.99	.08	-1.34
2	.72	.27	.19
3	.90	.21	.11
4	.70	.07	5.87
5	.70	.32	-1.45
6	.50	.32	-1.64
7	.84	.26	-.40
8	.85	.25	-.31
9	.89	.04	2.13
10	.86	.35	-2.05
11	.98	.20	-.81
12	.81	.29	-.73
13	.91	.23	-.42
14	.57	.39	-4.56
15	.32	.28	-1.47
16	.88	.33	-1.50
17	.47	.22	2.20
18	.20	.21	.03
19	.80	.28	-.51
20	.53	.26	.62
21	.57	.17	4.51
22	.15	.16	.94
23	.66	.31	-1.13
24	.57	.30	-.90

Note. *N* = 1,350.

should be interpreted as figure and what as snow. With Item 9 it appears that many examinees may interpret some of the splatters as part of the fragmented figure. With Items 4 and 21, it appears that examinees may fail to include all segments of the figures in their closure. Furthermore, for each of these items, misinterpretation of a fragment of the picture as figure or snow is apt to result in closure of an identifiable albeit incorrect object rather than no closure at all. If this worksample were to be used by the Foundation in the future, modification of these pictures or the substitution of less ambiguous items should be considered. In addition, more-difficult replacement items may improve the reliability of Snowy Pictures, as well as provide better differentiation at the upper end of the score range.

A principal components factor analysis of the item correlation matrix yielded seven factors with eigenvalues greater than 1.0, namely, 3.18, 1.29, 1.14, 1.13, 1.10, 1.05, 1.01. The total variances attributable to the first and second factors were 13.2% and 5.4%, respectively. Each of the remaining factors accounted for less than 5% of the variance. As a result of the scree plot, it was concluded that a one-factor model was adequate to describe the item data.

Sex and age effects. There was neither a significant sex difference nor a significant age-by-sex interaction for Snowy Pictures. However, the analysis of variance showed a significant age effect for this worksample, $F(8;1,296) = 17.57$, $p < .01$. When scores are adjusted for sex, the proportion of variance accounted for by age alone is .10, which is comparable to the effect of a correlation of .32.

Figure 10 depicts the age curve for Snowy Pictures. Scores tend to increase with age until the mid-20s. Performance declines slightly from the mid-20s to the mid-30s, followed by rapid decline after age 35.

Summary. The findings related above with regard to Snowy Pictures can be summed up as follows:

1. The obtained split-half reliability coefficient of .61 was lower than desirable.

2. The average item difficulty was .68.

3. Approximately 30% of the items correlated at least .30 with total score.

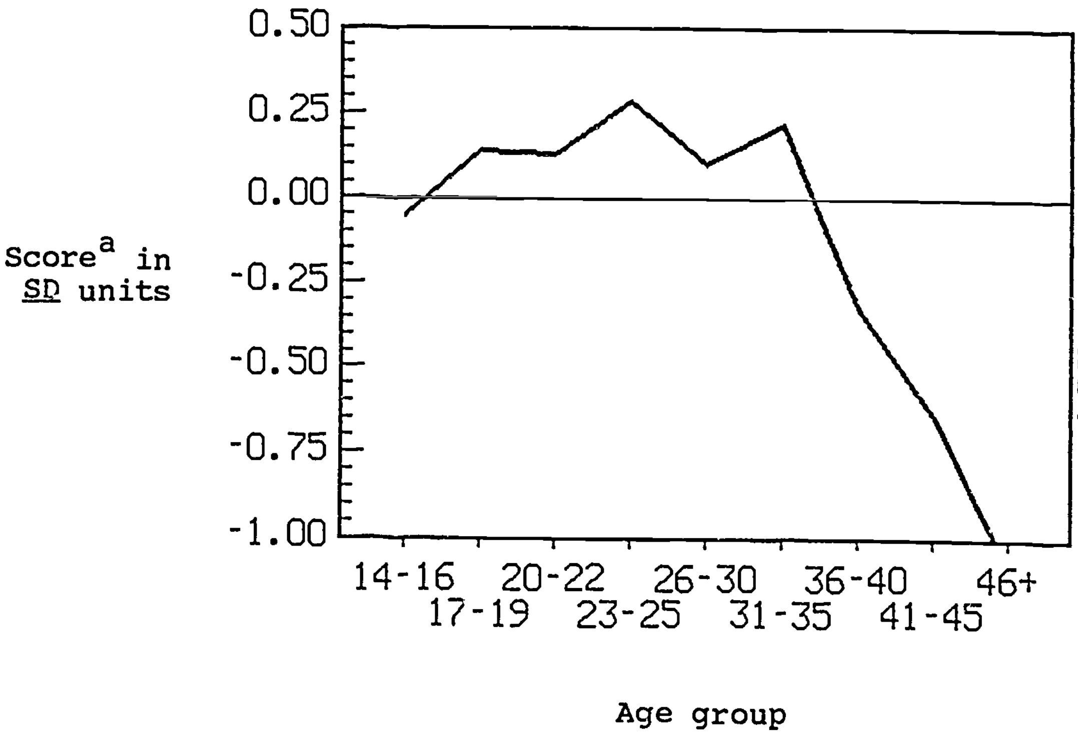
4. Three of the items were misfitting (i.e., were poor measures of the latent trait).

5. The reliability and item-analysis data suggest that Snowy Pictures either is not a good measure of the underlying construct or contains items that need improvement.

6. The age effect for this worksample was significant. There was no significant difference in performance between the sexes.

Figure 10

Age Curve for Snowy Pictures



Note. Plotted values are not smoothed.
[a]Adjusted for sex.

Concealed Words Test

Scores. As was indicated earlier, the Concealed Words Test was divided into two comparable parts. Examinees received a score for each part based on the number of items correctly completed. A score for the full test was calculated by adding examinees' two part scores. Table 6 shows descriptive statistics for each part of Concealed Words. There was no evidence of either a floor or ceiling effect. The distribution of total test scores approximated the normal curve, with scores spread widely across the entire range, except at the extreme ends.

Table 6

Descriptive Statistics for Concealed Words Test

Section	No. of items	Mean	SD	Range
Part 1	25	12.7	3.7	1-22
Part 2	25	13.2	4.5	1-25
Total	50	26.0	7.5	5-46

Note. *N* = 1,261.

Internal structure. A split-half correlation of .65 was calculated for the Concealed Words Test. Applying the Spearman-Brown correction to the correlation resulted in a reliability estimate of .79, which is near the Foundation's recommended minimum standard of .80.

Table 7 displays the item difficulties, item-total correlations, and infit statistics for the test. Item difficulties ranged from a low of .01 (for Item 20) to a high of .99 (for Item 3). The average difficulty for the worksample was .52, with a good distribution of *p* values across the entire test. Items 1 and 3 from Part 1 were very easy for the Foundation sample. Although they provided no discrimination of person ability, they may serve as good warm-up items. Four of the five most difficult items (*p* values less than .10) were from Part 1. Item 20, in particular, was extremely difficult for the Foundation sample; only 1% of the examinees answered that item correctly.

Item-total correlations ranged from a low of .06 for the two easiest items to a high of .53. Five items from Part 1 (1, 3, 20, 22, and 23) and two items from Part 2 (30 and 45) displayed low item-total correlations. Twenty-nine of the 50 items (11 from Part 1 and 182 from Part 2) correlated at least .30 with total score. On the basis of item-total correlations, few of the items on the Concealed Words Test can be regarded as exceptionally good. Nevertheless, when reliability relative to test administration time is considered, Concealed Words appears to be a more efficient worksample than either Gestalt Completion or Snowy Pictures.

Infit statistics ranged from -7.60 to 6.14. Items 21 and 45 appeared to be extremely misfitting. In Item 21, those sections of the letter "e" that are not concealed look very similar to parts of the letter "a." Consequently, 10% of the examinees

Table 7

Item Statistics for Concealed Words Test

Item	Difficulty (p value)	Item-total correlation	Fit statistic
1	.96	.06	.51
2	.80	.30	.14
3	.99	.06	-.58
4	.91	.25	.00
5	.82	.36	-1.43
6	.85	.27	.34
7	.49	.45	-3.73
8	.57	.31	2.15
9	.81	.33	-.45
10	.59	.30	2.[illegible]9
11	.63	.29	2.21
12	.39	.35	.58
13	.32	.39	-.94
14	.75	.22	2.99
15	.41	.48	-5.07
16	.36	.31	1.90
17	.24	.21	3.28
18	.24	.29	1.25
19	.05	.23	-.61
20	.01	.14	-2.67
21	.61	.24	4.73
22	.04	.15	-.18
23	.08	.10	1.14
24	.36	.27	3.16
25	.44	.39	-1.16
26	.79	.38	-1.80
27	.91	.24	-.10
28	.93	.20	-.27
29	.77	.24	1.94
30	.90	.17	.96
31	.78	.38	-1.82
32	.86	.26	.29
33	.58	.30	2.73
34	.43	.30	2.92
35	.65	.36	-.36
36	.33	.46	-3.97
37	.41	.48	-4.92
38	.43	.53	-7.60
39	.72	.35	.02
40	.63	.43	-2.81
41	.53	.39	-1.08
42	.19	.36	-1.58
43	.33	.53	-7.07
44	.05	.31	-1.56
45	.34	.16	6.14
46	.67	.33	.76
47	.10	.36	-1.70
48	.16	.31	-.61
49	.40	.45	-3.62
50	.37	.28	2.60

Note. *N* = 1,261.

closed the word as "woman" instead of "women." A reason for the poor item statistics for Item 45 is less obvious, except for the fact that the unconcealed portion of the last letter is rather small, which may make rapid closure difficult. Items 8, 10, 11, 14, 17, 24, 33, 34, and 50 showed moderately poor fit relative to the expected pattern of responses. The other items displayed average or better fit. With the exception of Item 20, the items that correlated more than .40 with total score were also the best items on the basis of the infit statistics.

The item factor analysis for the Concealed Words Test had to be handled in a special way. Because of computer limitations, the items were split into two subsets, odd-numbered and even-numbered, and a principal components factor analysis was run on each set of 25 items. For the odd-numbered items, factoring resulted in 7 factors with eigenvalues greater than 1.0, namely, 4.13, 1.34, 1.24, 1.10, 1.08, 1.04, and 1.00. The first factor accounted for 16.5% of the total variance, while the second accounted for 5.4%. None of the remaining factors accounted for more than 5% of the variance. For the even-numbered items, factor extraction likewise yielded 7 factors with eigenvalues greater than 1.0, namely, 3.68, 1.40, 1.27, 1.19, 1.09, 1.04, and 1.01. The total variance attributable to the first factor was 14.7%. The second and third factors accounted for 5.6% and 5.1%, respectively, of the variance. Each of the remaining factors accounted for less than 5% of the total variance. From the scree plots it appeared that a one-factor model was the most appropriate for both the odd-numbered items and the even-numbered items from the Concealed Words Test.

To ascertain if the first factor extracted for the odd-numbered items was equivalent to the first factor extracted for the even-numbered items, the two factor scores obtained for each examinee were correlated. The correlation between the odd and even factor scores was .78, which, given that the split-half correlation was .65, indicates very good agreement between the two factors. This means that, if scoring of the Concealed Words Test were based on the loadings for the first factor, the worksample's reliability would be .88.

<u>Sex and age effects</u>. For this worksample, a significant main effect for sex was found, $F(1;1,215) = 7.33$, $p < .01$. The analysis of variance also indicated the main effect for age was significant, $F(8;1,215) = 12.14$, $p < .01$. The sex-by-age interaction was not significant.

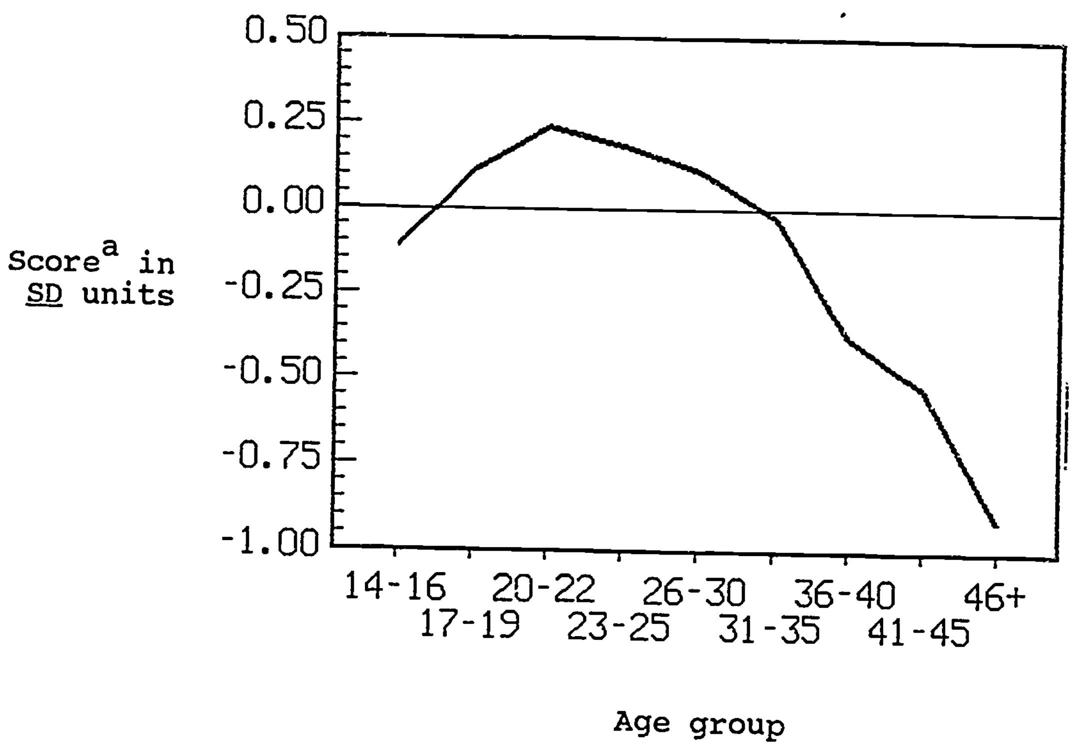
The difference between males and females in scores on Concealed Words is .15 of a standard deviation, with males performing slightly better than females (male $M = 26.70$, female $M = 25.23$). The proportion of variance accounted for by sex alone is negligible ($w^2 < .01$) and is comparable to less than the effect of a correlation of .10.

Approximately 7% of the variance associated with Concealed Words is attributable to age alone. This is equivalent to the effect of a correlation of .26.

The age curve for this worksample is presented graphically in Figure 11. Performance tends to increase with age until the early 20s and then declines slightly through the mid-30s, with sharper declines observed after age 35.

Figure 11

Age Curve for Concealed Words Test



Note. Plotted values are not smoothed.
[a]Adjusted for sex.

Summary. The above-described findings can be summarized as follows:

1. The Concealed Words Test is moderately reliable, possessing a split-half reliability coefficient of .79.

2. The average item difficulty of .52 was in the desirable range, and there was a well-balanced distribution of item difficulties.

3. Eighteen of the 50 items correlated at least .35 with total score.

4. Infit statistics revealed that two items have extremely poor fit.

5. Both age and sex effects were encountered in the worksample, but the sex effect was quite small.

Hidden Patterns Test

Scores. As noted earlier, the Hidden Patterns Test was divided into two parallel parts. Attempted items within each part were scored as to whether or not they were answered correctly. An examinee's score for a particular part consisted of the number of items he or she answered correctly within the time limit, corrected for guessing by subtracting the number of items answered incorrectly. A score for the entire test was computed by adding together an examinee's part scores. Descriptive statistics for the test are given in Table 8.

Table 8

Descriptive Statistics for Hidden Patterns Test

Section	Total correct	SD	Range[a]	Total attempted	SD	Range[a]
Part 1	100.2	23.0	10-167	101.9	22.4	16-169
Part 2	100.4	19.6	22-159	103.0	19.5	31-165
Total	200.6	40.7	44-322	204.9	39.7	50-329

Note. N = 1,276.
[a]Number of items in each part is 200, for a total of 400.

There were no floor or ceiling effects. The distribution of scores approximated the normal curve, with scores distributed across the range, except at both extremes. The number of items answered incorrectly ranged from 0 to 82. Total scores, corrected for guessing, ranged from 18 to 315.

Internal structure. As can be seen from Table 8, on the average, the differences between the number of items correct and the number of items attempted were very small (1.7, 2.6, and 4.3 for Part 1, Part 2, and the complete test, respectively). Overall, almost 98% of the items attempted were answered correctly.

As previously mentioned in the Method section, when most of the items attempted are answered correctly, there is not sufficient variance at the item level to conduct meaningful item-level analyses. Consequently, the internal structure of Hidden Patterns could not be analyzed in the usual manner.

Two estimates of reliability were calculated for the worksample. When scores were not corrected for guessing, the correlation between Parts 1 and 2 of the Hidden Patterns Test was .82, yielding a reliability coefficient of .90. When scores were corrected for guessing, the correlation between the two parts was .84, yielding a reliability coefficient of .91. Although high reliability generally is indicative of a test's unidimensionality, it is possible for a test to have a high reliability coefficient without being unidimensional. Normally an item factor analysis would have been conducted to confirm the unidimensionality of the Hidden Patterns Test, but this could not be done because of the lack of variance at the item level. However, it can probably be assumed that the test is essentially unidimensional.

Sex and age effects. Males and females did not differ significantly in their performance on the complete worksample. There was, however, a significant age effect, $F(8;1,230) = 3.21$, $p = .001$. When scores are adjusted for sex, the proportion of variance accounted for by age alone is relatively small ($w^2 = .02$), which is comparable to the effect of a correlation of .14. The interaction between sex and age was not significant.

Figure 12 depicts the age curve for Hidden Patterns. The adult plateau appears to range from the mid-20s through the mid-30s, after which scores decline modestly and then level off.

Summary. The main findings with regard to the Hidden Patterns Test can be summarized as follows:

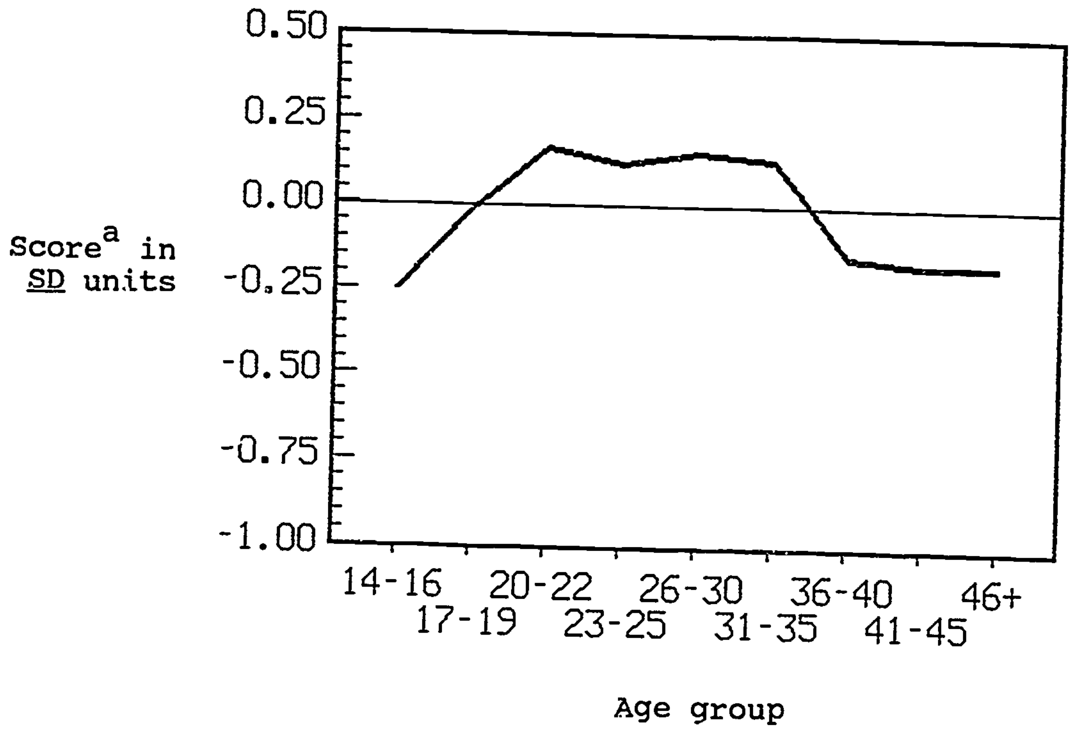
1. The worksample is highly reliable, possessing a split-half reliability coefficient of .91 when scores are corrected for guessing.

2. Although verification of the test's unidimensionality through item factor analysis was not possible, the worksample appears to measure only one dimension.

3. A significant main effect for age was encountered, but no sex difference was found.

Figure 12

Age Curve for Hidden Patterns Test



Note. Plotted values are not smoothed.
[a]Adjusted for sex.

Hidden Figures Test

Scores. As was indicated earlier, only one of the two parts of the Hidden Figures Test was administered. Attempted items were scored as to whether or not they were answered correctly. An examinee's score for the test was the number of items answered correctly minus a guessing factor of one-fourth the number of items answered incorrectly. Descriptive statistics for the test are given in Table 9.

Total scores on the worksample were spread across the entire range, with a steep incline at the low end of the curve and a piling up of scores at the test ceiling. Nevertheless, the distribution was sufficiently normal for statistical analyses.

As can be seen from Table 9, on the average, the difference between the number of items correct and the number of items attempted was 1.8 items. Overall, 80% of the items attempted were answered correctly.

Table 9

Descriptive Statistics for Hidden Figures Test

Part 1[a]	Mean	SD	Range
No. attempted	9.0	4.0	1 - 16
No. correct	7.2	4.1	0 - 16
Total score	6.7	4.4	-4 - 16

Note. N = 1,343.

[a]Part 2 of the test was not administered. Number of items in Part 1 is 16.

Internal structure. The procedures outlined in the Method section for deriving reliability estimates for speeded tests (i.e., using time-based subtest scores) could not be applied to Hidden Figures, because only Part 1 of the worksample was administered. Internal-consistency indices, albeit not appropriate for speeded tests, nonetheless provide a measure of the upper limit of a test's reliability. Consequently, coefficient alpha was computed as the estimate of reliability for Hidden Figures, although this was expected to result in a spuriously high estimate. When the variances used in the calculation were based on scores not corrected for guessing, an upper-bound alpha of .84 was obtained for the Hidden Figures Test; when the variances were based on scores corrected for guessing, an upper-bound alpha of .85 resulted.

Another reliability estimate for Part 1 only of the Hidden Figures test was calculated by utilizing reported reliabilities for the whole test (specifically, .80 for females and .82 for males; Ekstrom, French, Harman, & Derman, 1976) and applying the Spearman-Brown formula for reliability of a test of length .5, that is, a test of half the length. The resulting reliability, an estimate for Part 1 only of Hidden Figures, was .68. To increase the reliability to an acceptable .80 would require almost doubling the length of the worksample (from 16 items to 30 items), which would increase total test time from 12 to 23 minutes. In essence, this means administering the full test rather than Part 1 only.

Table 10 displays the item difficulties, item-total correlations, and infit statistics for Part 1 of the worksample. Item difficulties ranged from .11 for Item 15 to .74 for Item 1, with an average difficulty of .45. None of the items was extremely difficult or simple. The evidence suggests that the distribution of item difficulties does not encompass the entire

Table 10

Item Statistics for Hi en Figures Test

Item	Difficulty (p value)	Item-total correlation	Fit statistic
1	.74	.32	3.68
2	.66	.45	-.76
3	.50	.45	2.53
4	.69	.51	-3.71
5	.52	.48	-.04
6	.51	.49	-.00
7	.48	.53	-2.07
8	.38	.50	-.79
9	.46	.58	-5.63
10	.35	.52	-2.18
11	.53	.45	.65
12	.40	.57	-4.94
13	.30	.50	-1.92
14	.41	.33	6.39
15	.11	.29	1.28
16	.13	.39	.09

Note. *N* = 1,343.

range of person ability. For Part 1 to stand alone as a test, several easier items appear to be needed along with at least one more difficult item.

Item-total correlations ranged from .29 to .58, with none of the items displaying seriously low item-total correlations. Item 15 correlated .29 with total score; the remaining 15 items correlated at least .32 with total score. Infit statistics ranged from −5.63 to 6.39. Item 14 exhibited extremely poor fit, while Items 1 and 3 showed moderately poor fit relative to the expected pattern of responses. The other items displayed at least average fit.

Principal components factoring of the item correlation matrix produced three factors with eigenvalues greater than 1.0, namely, 4.89, 1.87, and 1.02. The first factor accounted for 30.5% of the variance in the test, while the second accounted for 11.7%. Each of the remaining factors accounted for less than 7% of the variance. One- and two-factor solutions were examined. The two-factor solution did not appear to divide the items on substantive grounds; rather, they were divided by difficulty and position in the test. It was therefore concluded that a one-factor model was appropriate for the Hidden Figures Test.

<u>Sex and age effects</u>. There was neither a significant sex difference nor a significant age-by-sex interaction for this worksample. However, the analysis of variance showed a significant age effect, $\underline{F}(8;1,289) = 2.24$, $\underline{p} < .025$. When scores are adjusted for sex, the proportion of variance accounted for by age alone is small ($\underline{w}^2 = .01$), which is comparable to the effect of a correlation of .10.

The age curve for Hidden Figures is depicted in Figure 13. Performance tends to increase with age until the early 20s, but the extent of the adult plateau is difficult to establish with the current sample. Although it appears that performance may begin to decline gradually after age 40, the differences in scores for examinees aged 20 through 46 and older are slight (range = −.04 to .16 of a standard deviation about the mean) and may be due to chance.

<u>Summary</u>. The findings related above with regard to the Hidden Figures Test can be summed up as follows:

1. Neither a split-half nor an internal-consistency reliability measure was appropriate for this test as administered in this study. The obtained alpha coefficient of .85 is likely to be spuriously high and merely indicates the upper limit of the worksample's reliability. Based on ETS's reported reliabilities for the full test, a reliability of .68 was calculated for Part 1 only.

2. The Hidden Figures Test possesses an average item difficulty of .45, but there is evidence that the distribution of difficulties does not encompass the range of person ability.

3. All but one of the item-total correlations were at least .32.

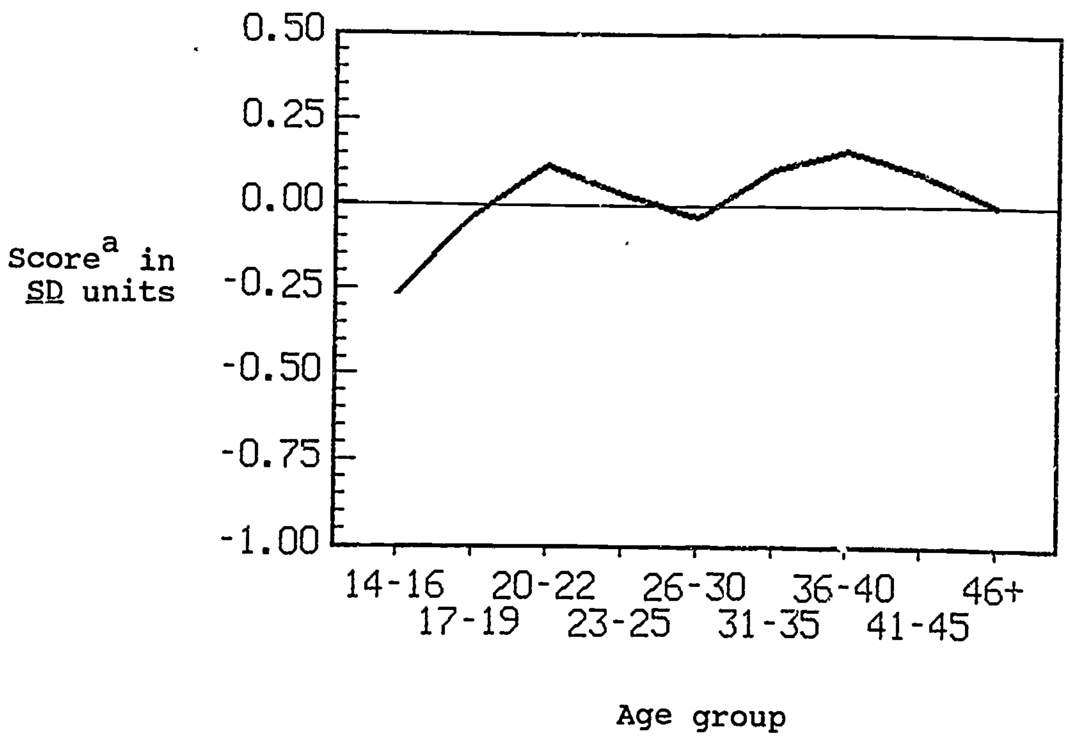
4. Infit statistics revealed one item with extremely poor fit.

5. The distribution of test scores and relatively restricted range of item difficulties suggest that when only Part 1 is administered, the worksample has too low a ceiling to discriminate properly at the higher end of the ability range. Furthermore, several easy items may be needed to balance the range of item difficulties.

6. The age effect for this worksample was significant, but there was no significant difference in performance between the sexes.

Figure 13

Age Curve for Hidden Figures Test



Note. Plotted values are not smoothed.
[a]Adjusted for sex.

Drawing

Scores. Each item of the Drawing test was scored based on the number of lines drawn that started and ended on the correct dots. In calculating each item score, the reference line was also counted as a line drawn correctly, so that the minimum score on each item was one rather than zero. A score for the whole test was obtained by totaling the examinee's scores on the ten items. An alternative total score was calculated by excluding from the total score the two items for which solutions were shown to the examinee and summing the examinee's scores on the other eight items. Table 11 presents the means and standard deviations for the items in the Drawing worksample.

Scores on the entire test (M = 36.95, SD = 17.60) ranged from 12 (two points above the minimum) to the maximum of 89. When Items 1 and 2 were excluded, scores ranged from the test floor (i.e., 8) to the test ceiling (i.e., 72). The distribution curve was clearly skewed, with scores concentrated toward the low end of the range. There were no appreciable floor or ceiling effects.

Internal structure. Coefficient alpha was computed as the estimate of reliability for Drawing. When all ten items were included in the total score, an alpha of .87 was obtained; when total score included only Items 3 through 10, an alpha of .88 resulted. From the reliability analysis neither total score emerged as clearly superior. Therefore, all ten items were included in the internal-structure analyses for this worksample.

As can be observed from Table 11, Items 1, 2, and 10 proved to be the easiest, while Items 5 and 9 were the most difficult. For this sample, the distribution of mean scores appears much too restricted for the range of person abilities. In particular, the worksample cannot properly differentiate examinees at the lower end of the ability range. Including some easier items (e.g., drawings with fewer than nine lines) might help to balance the range of item difficulties.

Table 11 presents the percentage of examinees who failed to draw any of the lines correctly for an item. Fewer than 5% of the examinees received the minimum score for Items 2 or 3, while approximately half of the examinees tested failed to draw correctly any of the lines in Items 5 or 9. Based on either mean score or percentage of examinees receiving the minimum score, Items 5 and 9 seem to be the most difficult. This is probably due to the slopes of the lines contained in each drawing. Horizontal and vertical lines (i.e., those with 0° and 90° slopes, respectively) appear to be the easiest to reproduce, followed by lines with 45° slopes. There is also evidence that the more the slope of a line deviates from 45° the more difficult it is to draw correctly. Thus, a line with a slope of 10° (which deviates 35° from a 45° slope) is more difficult than a line with a slope of 70° (which deviates 25°). From this perspective,

Table 11

Item Statistics for Drawing

Item	Mean[a]	SD	% with minimum score	Item-total correlation
1	5.34	3.22	17.2	.32
2	4.11	2.31	2.4	.48
3	3.58	2.14	4.0	.67
4	3.32	2.49	30.1	.64
5	2.57	2.28	51.3	.68
6	3.26	2.63	39.8	.66
7	3.21	2.60	38.4	.52
8	3.91	3.01	31.0	.71
9	2.93	2.67	47.8	.66
10	4.74	2.40	14.8	.65

Note. *N* = 1,292.

[a]The minimum score for each item was 1. The maximum score was 8 for Item 2 and 9 for Items 1 and 3 through 10.

Items 5 and 9 contain more lines with slopes that deviate greatly from 45° than the other items on the Drawing worksample.

Corrected item-total correlations for each item in the worksample are also contained in Table 11. They ranged from .32 to .71. None of the items displayed seriously low item-total correlations. Only Item 1 had a moderately low correlation with total score.

Coefficient alphas were recalculated for the worksample, with Items 1 and 2 excluded in turn from the analysis. Omitting Item 1 only did not result in any change in the internal-consistency reliability of Drawing, but omitting Item 2 only resulted in a slightly lower coefficient alpha of .865. It appears, then, that including Item 2 in the worksample enhanced reliability, whereas including Item 1, with its moderately low item-total correlation, neither enhanced nor detracted from the test's reliability.

Because of the high alpha coefficient of the Drawing test, it was expected that this worksample would prove to be essentially unidimensional. This was confirmed by factor analysis. Principal components factoring yielded only one factor with an eigenvalue greater than 1.0, namely 4.87. The proportion of total variance attributable to this factor was 48.7%. This indicated that a one-factor model was optimal for describing the item data for this worksample and that all items are measuring the same underlying trait.

Most Foundation worksamples include at least one practice item that is not scored. Consequently, in the remainder of the analyses for this study, Items 1 and 2 of Drawing were treated as practice items and only Items 3 through 10 were used in computing total score. Because the worksample's reliabilities are so similar whether eight, nine, or 10 items are scored, it is recommended that in any future use of the test by the Foundation the first, if not both, of the practice items be excluded from the calculation of total score.

Sex and age effects. The analysis of variance of scores by sex and age group showed a significant sex effect, $F(1;1,239) = 97.04$, $p < .001$. Neither an age effect nor a sex-by-age interaction was found.

The mean score for males on this worksample was 31.59, and the mean score for females was 23.29. When scores are adjusted for age, the difference in performance between males and females is .54 of a standard deviation. The proportion of variance in Drawing that is accounted for by sex alone is .07, which is comparable to the effect of a correlation of .26.

Although the age main effect was not significant, scores on Drawing appear to increase modestly with age until the early 20s and then level off through the mid-20s. But after the mid-20s, the shape of the age curve appears unstable, so that further interpretation would not be fruitful.

Summary. The above-described findings can be summarized as follows:

1. The Drawing test is moderately reliable, possessing an alpha of .88.

2. The skewed distribution of test scores and the restricted range of item means suggest that the worksample probably contains too many difficult items for the Foundation sample.

3. Item 1 correlated .32 with total score. All the other item-total correlations were at least .48.

4. Males outperformed females by approximately one-half of a standard deviation. The main effect for age was not significant.

5. If the worksample is used in the future, it is recommended that at least the first item be considered a practice item and therefore excluded from the calculation of total score.

Maze Tracing Speed Test

Scores. As previously mentioned, the Maze Tracing Speed Test was divided into two parallel parts. The number of mazes through which a path was correctly drawn within a part determines an examinee's score for that part. A score for the whole test was obtained by totaling an examinee's scores on Parts 1 and 2. Descriptive statistics for each part of Maze Tracing are presented in Table 12.

Table 12

Descriptive Statistics for Maze Tracing Speed Test

Section	No. of items	Mean	SD	Range
Part 1	24	12.9	4.1	2-24
Part 2	24	15.0	4.2	2-24
Total	48	28.0	8.0	6-48

Note. N = 1,351.

Scores on both Part 1 and Part 2 ranged from 2 to 24. Total scores on the worksample ranged from 6 to the maximum of 48, with some indication of a ceiling effect. Except for a piling up of scores at the test ceiling, the distribution approached a normal curve, with scores spread widely enough across the range for differences among examinees to exhibit themselves.

Internal structure. Because the Maze Tracing Speed Test is almost a pure speed measure (i.e., most, if not all, of the mazes attempted are completed correctly), there is not sufficient variance at the item level to conduct meaningful item-level analyses. Consequently, because the statistics generally utilized in this study to assess the quality of a test's items (i.e., item difficulties, item-total correlations, Rasch fit statistics, and item factor analysis) were not appropriate, the internal structure of this worksample could not be analyzed in the usual manner.

The split-half correlation between Parts 1 and 2 of Maze Tracing was .86, yielding a reliability coefficient of .92. As explained earlier, a high estimate of reliability generally means a test is essentially unidimensional. Although it was not possible to confirm the worksample's unidimensionality by conducting an item factor analysis, for most purposes it can probably be assumed that the test measures only one dimension.

Sex and age effects. The analysis of variance identified a significant main effect for sex for this worksample, $F(1;1,350) = 8.04$, $p < .01$. The main effect for age was also significant, $F(8;1,350) = 9.07$, $p < .01$. The sex-by-age interaction was not significant.

Males completed an average of 32.30 mazes, while females completed an average of 29.25. When scores are adjusted for age, this represents a difference in performance between the sexes of .15 of a standard deviation. The proportion of variance accounted for by sex alone is negligible ($w^2 < .01$) and is comparable to less than the effect of a correlation of .10.

Approximately 5% of the variance associated with the Maze Tracing Speed Test is attributable to age alone. This is equivalent to the effect of a correlation of .22.

The age curve for Maze Tracing is depicted in Figure 14. Performance on this worksample tends to decrease with age. Whether the absence of a growth curve is true of the spatial-scanning factor in general or specific to the Maze Tracing Speed Test, however, is an aspect of perceptual abilities beyond this study's purview.

Summary. The main findings with regard to the Maze Tracing Speed Test can be summarized as follows:

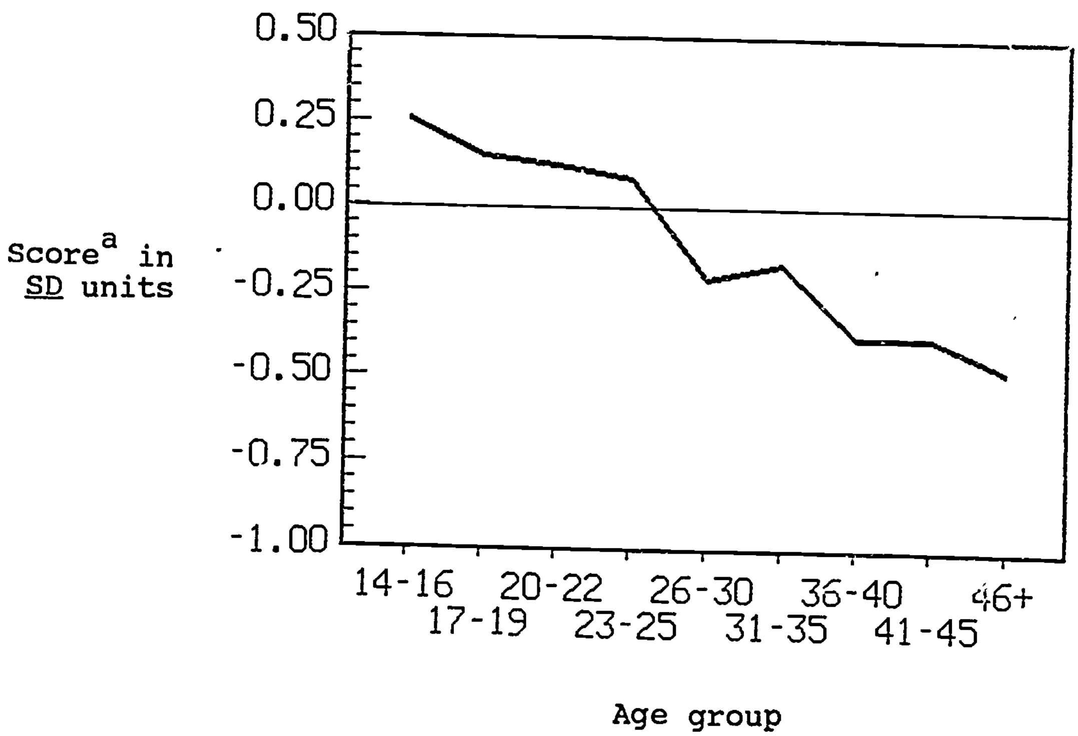
1. With a split-half reliability coefficient of .92, the worksample is highly reliable.

2. Although an item factor analysis could not be conducted, it can probably be assumed that the test is essentially unidimensional.

3. Main effects for both sex and age were encountered, but the sex effect was quite small. Scores on Maze Tracing tend to decline with age.

Figure 14

Age Curve for Maze Tracing Speed Test



Note. Plotted values are not smoothed.
[a]Adjusted for sex.

Tactics

Scores. An examinee's score for each item was the time it took the examinee to complete that trial correctly. An answer was scored as correct if it was the most efficient route in terms of the number of turns taken between the "Start" and "Finish" squares. If an examinee offered a solution that did not follow the rules, he or she was instructed to continue working. No penalty was assessed for such impermissable answers. If an examinee failed to solve an item within two minutes, he or she was assigned a score of 2.00 for that trial. A score for the full worksample was calculated by adding together the examinee's item scores.

As mentioned in the Method section, each examinee was administered either a five-trial or a 10-trial version of

Tactics, with Items 1 through 5 in the longer version the same as Items 1 through 5 in the shorter version. A comparison of the five-trial-only group (n = 815) and the 10-trial group (n = 445) revealed no significant difference in performance on Items 1 through 5, $t(1,258) = .71$, $p = .48$. Consequently, the item-level data for Trials 1 through 5 are based on item scores from the entire sample, while the data for Trials 6 through 10 are based on the scores of a subset of the sample.

Table 13 shows the mean and standard deviation for each item of the Tactics test. Total test times ranged from .4 minutes to approximately 8 minutes on the five-item version and from 1.5 to almost 13 minutes on the 10-item version. Scores approached a normal distribution about the mean and were spread widely enough across the range for differences among examinees to exhibit themselves. There was no indication of either a floor or a ceiling effect.

When the scoring of a test is based on time to solution rather than number completed correctly within a given amount of time, a nonlinear transformation of the time-based score is psychometrically appropriate (Cohen & Cohen, 1983). For the Tactics test, the transformation was accomplished by taking the reciprocal of the time for each item. An examinee's transformed total score was the sum of his or her transformed item scores. Transformed scores were utilized for most of the analyses involving Tactics.

Internal structure. As can be observed from Table 13, mean time to solution ranged from 10.7 to 168.6 hundredths of minutes. Trials 1, 3, and 9 showed the fastest times to completion, while Trials 5, 8, and 10 manifested the slowest times.

The percentage of examinees who solved an item in the allotted time period is also contained in Table 13, along with corrected item-total correlations for each item in the worksample. Six of the items (1, 2, 3, 6, 7, and 9) were solved within the two-minute time period by almost all of the examinees. Trial 8 proved to be the most difficult to solve, with only 30% of the examinees completing it correctly within the allotted time. Item-total correlations based on transformed scores ranged from a low of .13 to a high of .55. Trials 5, 8, and 10 displayed low item-total correlations. The remaining items correlated at least .35 with transformed total score. Trials 5, 8, and 10 do not correlate with total test score and also do not correlate with each other or any of the other Tactics items.

Coefficient alpha was computed as the estimate of reliability for Tactics. When the variances used in the calculation were based on transformed scores, an alpha of .73 was obtained, which is lower than desirable by Foundation standards. To increase the reliability to an adequate .80 would necessitate lengthening the

Table 13

Item Statistics for Tactics

Trial	Mean time[a]	SD	Percentage with time < 2.00[b]	Item-total correlation[c]
1	17.6	20.1	.99	.53
2	26.8	32.4	.99	.47
3	16.7	18.9	.99	.55
4	69.0	63.6	.88	.45
5	134.6	66.8	.61	.13
6	31.4	33.3	.99	.49
7	30.4	29.1	.99	.36
8	168.6	56.1	.30	.19
9	10.7	13.1	.99	.52
10	146.8	59.9	.55	.17

Note. *N* = 1,260 for Trials 1 through 5; *N* = 445 for Trials 6 through 10.

[a] In hundredths of minutes.
[b] Percentage with time < 2.00 minutes is the percentage of examinees who solved the item within the time limit.
[c] Based on transformed scores.

worksample to 15 items. This would result in a test lasting 11 minutes, on the average, although in some cases, the longer version of Tactics could take up to 30 minutes to complete. Coefficient alphas were recalculated for the worksample, with individual items excluded in turn from the analysis. Omitting Trial 5, 8, or 10 did not result in any change in the internal-consistency reliability of Tactics. In other words, even though Trials 5, 8, and 10 have low item-total correlations, including them in the worksample did not detract from the test's reliability, although replacing those trials with better items might improve the reliability to a nearly acceptable .79.

One feature that distinguishes Trials 5, 8, and 10 from the other items is that, for the former, the shortest routes in terms of physical distance between the "Start" and "Finish" squares are not the most efficient routes in terms of the number of turns required. Figure 15 illustrates the difference between short and efficient routes. It is probable that the underlying construct measured by items in which the shortest route is also the most efficient route (perceptually consistent items) is different from the underlying construct measured by the items in which the shortest route is not the most efficient (perceptually inconsistent). This would explain the lack of fit of Trials 5, 8, and 10 on the worksample. Before being used further by the Foundation, the Tactics test should be revised so that all the items are perceptually consistent.

Principal components factoring of the item correlation matrix resulted in two factors with eigenvalues greater than 1.0, namely 3.02 and 1.06. The first factor accounted for 30.2% of the variance in the test, while the second factor accounted for 10.6% of the variance. None of the remaining factors accounted for more than 10% of the variance. From the scree plot it appeared that a one-factor model was the most appropriate for the Tactics test.

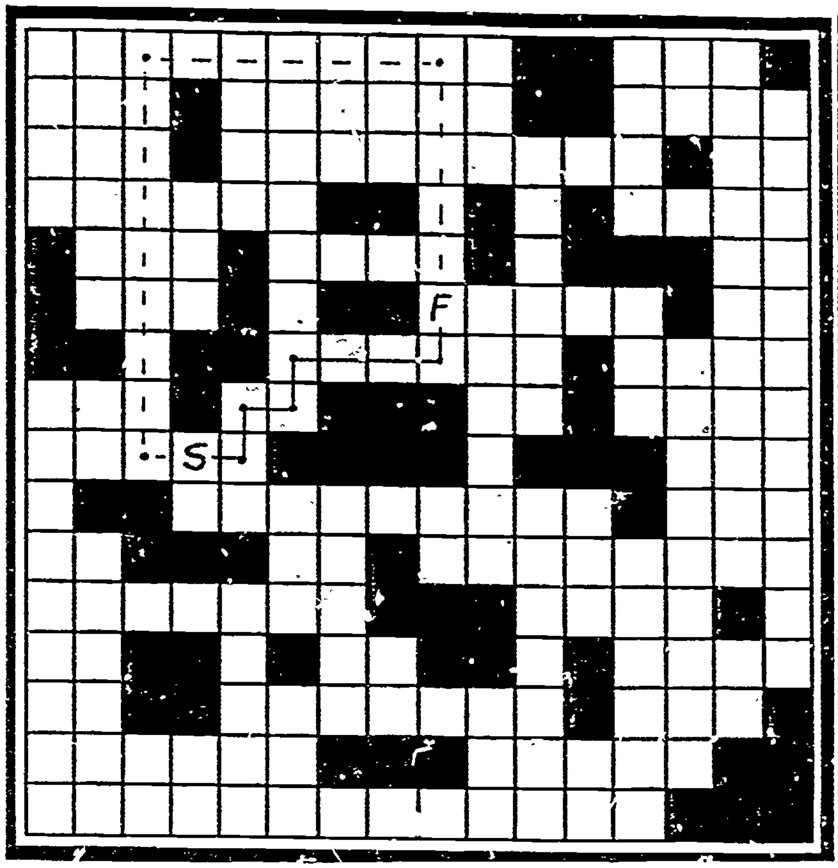
<u>Sex and age effects</u>. Inasmuch as the findings regarding sex and age effects on performance on Tactics are similar for both the five-item and 10-item worksamples, results of the analyses are reported for only the longer version. The analysis of variance showed significant main effects for sex and age: $F(1,421) = 48.48$, $p < .001$, and $F(8,421) = 2.27$, $p < .03$, respectively. The sex-by-age interaction was not significant.

The difference in total transformed time between the sexes was .63 of a standard deviation, with males taking less time than females to complete the worksample. Approximately 10% of the variance associated with Tactics is attributable to sex alone. This is comparable to the effect of a correlation of .32.

When scores are adjusted for sex, the proportion of variance accounted for by age is .04. This is equivalent to the effect of a correlation of .20.

Figure 15

Example of a Perceptually Inconsistent Tactics Item



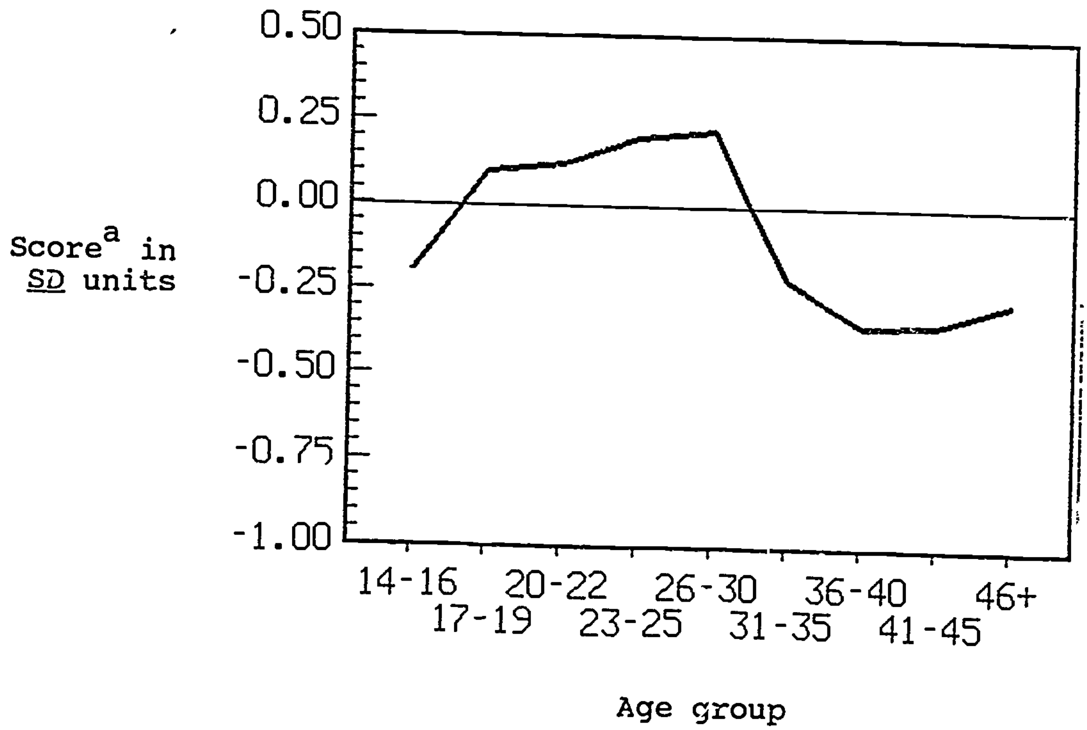
——— Solid line represents the shortest route, which requires five turns.

------ Dotted line represents the most efficient route, which requires only three turns.

Figure 16 depicts the age curve for the 10-item Tactics test. Scores appear to increase with age until the late 20s, followed by a sharp decline. From age 36 on, there is no evidence of a decrement in performance.

Figure 16

Age Curve for Tactics



Note. Plotted values are not smoothed.
[a]Adjusted for sex.

Summary. The findings related above with regard to the Tactics test can be summed up as follows:

1. The Tactics test possesses a close-to-acceptable alpha reliability of .73.

2. Seven of the 10 trials correlated at least .35 with total score.

3. The reliability data suggest that the construct measured is not well measured by the Tactics worksample. Replacing three of the trials with perceptually consistent items might increase the test's reliability to .79, which is near the Foundation's recommended minimum of .80.

4. The worksample showed both age and sex main effects. Males outperformed females by almost two-thirds of a standard deviation.

Summary of the Experimental Worksamples Considered Individually

Table 14 contains the estimates of reliability computed for each of the worksamples in the perceptual abilities battery. The reliabilities of Hidden Patterns, Drawing, and Maze Tracing fall in the acceptable range (.80 or higher). Concealed Words, with a reliability of .79, can also be considered acceptable. Tactics has a close-to-adequate reliability, but the worksample would need some refinement before it could be incorporated into the standard battery. Gestalt Completion, Snowy Pictures, and Hidden Figures (Part 1) have lower-than-desirable reliabilities and require improvement before being used further by the Foundation.

Because it is expedient to assess the experimental tests on the basis of practical considerations, such as ease and speed of administration, as well as psychometric properties such as reliability, the length of each test as administered in this study is also listed in Table 14. Moreover, for each worksample with a lower-than-desirable reliability, an estimate of the length needed for the test to reach an acceptable reliability of .80 is reported. Administration times for the reliable tests range from six to 10 minutes, so that in terms of speed of administration, any of those worksamples could reasonably be added to the JOCRF battery. Gestalt Completion, to be reliable, would need to be lengthened to approximately eight minutes. It is estimated that a reliable version of Tactics would take, on the average, 11 minutes to finish, although in some cases, the worksample could take up to 30 minutes to complete. Reliable versions of Snowy Pictures and Hidden Figures would require 21 and 23 minutes, respectively, to administer.

Sex and age effects for the individual worksamples are also included in Table 14. Three of the markers (Snowy Pictures, Hidden Patterns, and Hidden Figures) showed no significant sex effect. In addition, the sex effects for Gestalt Completion, Concealed Words, and Maze Tracing were negligible. For Drawing and Tactics, the two Foundation tests in the experimental battery, the difference in performance between males and females was one-half of a standard deviation or more.

All but the Drawing test exhibited an age effect. The smallest age effects were for the flexibility-of-closure measures, followed by the spatial-scanning worksamples and Gestalt Completion. Snowy Pictures and Concealed Words showed the largest age effects.

For Gestalt Completion, the adult plateau extended to age 30, whereas for Snowy Pictures and Concealed Words, scores increased until the early 20s, with a slight decline from the mid-20s to

Table 14

Summary Data for the Perceptual Abilities Worksamples

		Length of test		Length needed to increase rel. to .80			
Worksample	Rel.[a]	No. of items	Time[b]	No. of items	Time[b]	Sex effect[c,d]	Age effect[c]
Gestalt Completion	.68	20	4	38	8	.01	.03
Snowy Pictures	.61	24	8	62	21	—	.10
Concealed Words	.79	50	8	—	—	.01	.07
Hidden Patterns	.91	400	6	—	—	—	.02
Hidden Figures[e]	.68	16	12	30	23	—	.01
Drawing	.88	10	10	—	—	.07	—
Maze Tracing	.92	48	6	—	—	.01	.05
Tactics	.73	10	7	15	11	.10	.04

[a]Split-half reliabilities were computed for Gestalt Completion, Snowy Pictures, Concealed Words, Hidden Patterns, and Maze Tracing; alpha reliabilities were computed for Drawing and Tactics. Estimated reliability for Part 1 only of Hidden Figures was computed from reported reliabilites for the whole test of .80 for females and .82 for males (Ekstrom et al., 1976), using the Spearman-Brown formula for reliability of a test of length .5.

[b]In minutes, not including time for instructions.

[c]Reported in terms of the index $\underline{w}^2$. The square root of this index is equivalent to the effect of a correlation of the same value.

[d]For all worksamples with sex effects, males outperformed females.

[e]Data based on Part 1 only.

the mid-30s. After age 35, decreases in performance were observed for all three speed-of-closure markers.

Scores on Hidden Patterns, Hidden Figures, and Drawing increased through the early 20s. For Hidden Patterns, the adult plateau extended through the mid-30s, followed by a modest decline and then leveling off; for Hidden Figures and Drawing, the extent of the adult plateau and the beginning of decline was difficult to determine. The age curves for these flexibility-of-closure measures are in contrast to the age curve for structural visualization, the aptitude in the JOCRF battery with which these tests show their highest correlations (see later in report). Performance on structural visualization is at a maximum through the early 30s, after which time a gradual decline commences (see the Wiggly Block Worksample 3,4,5,6,7 Manual).

Scores on Tactics increased until the late 20s, followed by a sharp decline and then a leveling off from age 36. Unlike the other perceptual abilities worksamples, Maze Tracing showed no evidence of either a growth curve or an adult plateau.

In general, decreases in performance with age are larger for speeded tests than power tests. This was borne out in this study by the relatively modest decline observed for Hidden Figures, along with the more pronounced decreases seen for Gestalt Completion, Snowy Pictures, Concealed Words, and Maze Tracing.

The age curves for the perceptual abilities are consistent with patterns typically associated with aptitudes rather than acquired knowledge. In other words, they decline with age or maintain a relatively constant level during adulthood rather than increasing throughout adulthood, in the manner of vocabulary knowledge.

The Perceptual Abilities Battery

The zero-order correlations (simple Pearson product-moment coefficients) among the eight experimental worksamples are displayed in Table 15. The correlations among Gestalt Completion, Snowy Pictures, and Concealed Words ranged from .55 to .61, indicating relatively strong relationships among these speed-of-closure measures. Furthermore, the highest correlations for each of Gestalt Completion, Snowy Pictures, and Concealed Words were with the other two speed-of-closure markers, although Concealed Words also was moderately related to Hidden Patterns. Among the flexibility-of-closure worksamples (i.e., Hidden Patterns, Hidden Figures, and Drawing), the correlations ranged from .35 to .53. The highest correlations for Hidden Figures were with the other two measures of flexibility of closure. Both Hidden Patterns and Drawing correlated highest with Hidden Figures. However, the second highest correlation for Hidden Patterns was with Concealed Words; for Drawing, it was with Tactics. Tactics and the Maze Tracing Speed Test were only minimally related, suggesting that the two worksamples tend to

Table 15

Correlations Among the Perceptual Abilities Worksamples

Worksample[a]	GC	SP	CW	HP	HF	DR	MZ
Gestalt Completion							
Snowy Pictures	61 (39)						
Concealed Words	59 (43)	55 (38)					
Hidden Patterns	36 (28)	36 (27)	45 (38)				
Hidden Figures	29 (20)	25 (16)	40 (29)	53 (42)			
Drawing	25 (19)	19 (14)	36 (30)	35 (31)	50 (39)		
Maze Tracing	29 (23)	33 (25)	13 (11)	10 * (09)*	18 (14)	13 (12)	
Tactics	33 (23)	21 * (14)*	36 (27)	39 (32)	40 (28)	45 (36)	22 (18)

Note. Correlations are corrected for attenuation, with uncorrected correlations in parentheses. Leading decimals omitted. Ns range from 434 to 1,351. All correlations are significant at .001 level (2-tailed) unless otherwise indicated.

[a]Notation for worksamples: GC (Gestalt Completion); SP (Snowy Pictures); CW (Concealed Words); HP (Hidden Patterns); HF(Hidden Figures); DR (Drawing); MZ (Maze Tracing Speed Test).

*.001 < p < .01.

measure different traits. Both versions of Tactics correlated highest with the flexibility-of-closure measures. The Maze Tracing Speed Test was not highly related to any of the other worksamples in the perceptual abilities battery.

Initial principal components analysis with pairwise deletion extracted two factors with eigenvalues greater than unity. The first factor accounted for 35.8% of the total variance; the second factor accounted for 14.8% of the total variance. These two factors were then rotated to the varimax criterion. The resulting rotated factor matrix is presented in Table 16.

Five of the worksamples (Concealed Words, Hidden Patterns, Hidden Figures, Drawing, and Tactics) loaded (> .40) on the first rotated factor. Concealed Words also loaded on the second rotated factor, along with Gestalt Completion, Snowy Pictures, and Maze Tracing Speed Test. Although the three flexibility-of-closure markers loaded on Factor 1 and the three speed-of-closure markers loaded on Factor 2, one spatial-scanning worksample also loaded on Factor 1 and one on Factor 2. Distinct flexibility-of-closure, speed-of-closure, and spatial-scanning factors, as conceptualized by ETS, did not emerge from this analysis. Because it appeared that two factors might not be sufficient to represent the relationships among the experimental worksamples, an additional exploratory factor analysis was performed.

In this second analysis, three factors, which accounted for 63% of the variance, were rotated to the varimax criterion. The resulting rotated factor matrix for three factors is presented in Table 16. The first rotated factor exhibited high loadings for Hidden Patterns, Hidden Figures, Drawing, and Tactics. The second rotated factor revealed high loadings for Gestalt Completion, Snowy Pictures, Concealed Words, and, to a lesser degree, Hidden Patterns. Only one worksample, the Maze Tracing Speed Test, loaded on the third factor.

Although this factor solution bears similarities to the two-factor solution, the three-factor structure seems clearer in terms of interpretability of the factors. Factor 1 predominantly involves the flexibility-of-closure measures, and Factor 2 heavily involves the speed-of-closure worksamples, while Factor 3 is linked to a specific spatial-scanning test. Furthermore, the worksamples meant to mark each ETS-conceptualized factor load substantially more highly on the appropriate factor than on any other factor, with the exception of Tactics, which loads on the flexibility-of-closure factor.

It appears that a three-factor solution is more appropriate than a two-factor solution for representing the relationships among the eight perceptual abilities worksamples. In the three-factor structure, the flexibility-of-closure and speed-of-closure factors remained essentially intact, but a spatial-scanning factor marked by both the Maze Tracing Speed

Table 16

Factor Analyses of the Perceptual Abilities Worksamples[a]

Factor loadings for rotated[b] two-factor solution

	Factor 1	Factor 2
Gestalt Completion	.20	.74
Snowy Pictures	.08	.79
Concealed Words	.44	.56
Hidden Patterns	.66	.26
Hidden Figures	.74	.08
Drawing	.74	.04
Maze Tracing	.03	.54
Tactics	.64	.15

Factor loadings for rotated[b] three-factor solution

	Factor 1	Factor 2	Factor 3
Gestalt Completion	.13	.75	.17
Snowy Pictures	.01	.78	.20
Concealed Words	.34	.71	-.13
Hidden Patterns	.60	.41	-.17
Hidden Figures	.73	.15	-.02
Drawing	.75	.06	.08
Maze Tracing	.11	.17	.92
Tactics	.66	.08	.28

Note. Pairwise deletion of missing cases. *N*s range from 434 to 1,351.

[a]Initial extraction using principal components analysis.
[b]Varimax rotation.

Test and Tactics was not fully recovered. This was not totally unexpected, however, because a clear spatial-scanning factor has been reported in previous research only twice (Ekstrom et al., 1979). Factor 3 may be a spatial-scanning factor for which Maze Tracing is a marker, but this factor cannot be interpreted further because it is linked only to one specific worksample. Consequently, the extent to which the Maze Tracing factor represents a more-general perceptual ability such as spatial scanning cannot be determined without the inclusion of more tests in the factor analysis.

Although Tactics also has been thought to measure spatial scanning, the experimental worksample appears to correspond more closely to flexibility of closure. If the correlations of Tactics and the flexibility-of-closure tests with the JOCR standard worksamples are reasonably congruent, this would further substantiate a three-factor structure, with Tactics classified as a measure of flexibility of closure rather than spatial scanning. Investigation of the relationship between the experimental and standard worksamples was expected to shed additional light on this.

Summary. The main findings with regard to the relationships among the eight worksamples within the experimental perceptual abilities battery can be summarized as follows:

1. Gestalt Completion, Snowy Pictures, and Concealed Words were moderately correlated with each other, indicating relatively strong relationships among the three speed-of-closure markers. Hidden Figures correlated moderately with the other two flexibility-of-closure measures. The spatial-scanning tests were only minimally related. Tactics correlated highest with the flexibility-of-closur measures. The Maze Tracing Speed Test was not highly related to any of the other worksamples in the perceptual abilities battery.

2. In the exploratory factor analyses, the flexibility-of-closure, speed-of-closure, and spatial-scanning factors, as conceptualized by ETS, were not fully recovered. Nevertheless, the three-factor solution supported the *a priori* placement of the eight perceptual abilities worksamples on the speed-of-closure, flexibility-of-closure, and spatial-scanning factors, with the exception of Tactics, which loaded with flexibility-of-closure worksamples rather than the spatial-scanning test.

The Perceptual Abilities Battery and its Relationship to the JOCRF Standard Battery

Perceptual Abilities and the Cognitive Worksamples

Table 17 displays the correlations of the eight tests in the experimental battery with the cognitive worksamples in the standard battery. A number of moderate to high correlations were found between the perceptual abilities and the JOCRF worksamples.

Table 17

Correlations Between Perceptual Abilities Tests and Cognitive Worksamples in JOCRF Battery

Worksample[a]	GC	SP	CW	HP	HF	DR	MZ	TC	Rel.
Number Checking	ns	20 (15)	22 (19)	35 (33)	27 (22)	ns	10 (09)	ns	96
Ideaphoria	-10 (-08)	ns	ns	20 (19)	14 (11)	ns	ns	ns	97
Foresight	ns	ns	ns	14 (13)	12 (10)	ns	ns	18 (15)	97
Inductive Reas.	34 (26)	43 (31)	27 (22)	41 (36)	21 (16)	ns	13 (11)	20 (16)	84
Analytical Reas.	32 (21)	29 (18)	31 (22)	47 (36)	47 (31)	33 (25)	ns	44 (30)	65
Number Series	17 (13)	21 (15)	39 (32)	42 (37)	52 (40)	37 (32)	08 (07)	40 (32)	87
Wiggly Block	40 (28)	33 (22)	36 (27)	43 (35)	48 (34)	52 (42)	16 (13)	68 (50)	73
Paper Folding	46 (34)	33 (23)	43 (35)	43 (37)	58 (43)	59 (50)	16 (14)	56 (43)	82
Memory for Design	45 (33)	47 (33)	43 (34)	43 (37)	52 (38)	41 (34)	20 (17)	47 (36)	80
Silograms	13 (10)	20 (15)	23 (20)	19 (17)	19 (15)	ns	ns	13 (11)	92
Number Memory	17 (13)	28 (21)	29 (25)	26 (24)	27 (21)	22 (20)	09 (08)	26 (21)	91
Observation	18 (12)	33 (20)	16 (11)	29 (22)	23 (15)	11 (08)	11 (08)	12 (08)	62
English Vocab.	ns	ns	09 (08)	18 (17)	32 (26)	14 (13)	-15 (-14)	ns	96
Math Vocabulary	— (16)	— (09)	— (24)	— (28)	— (36)	— (31)	—	— (28)	—
Reading Effic.	13 (09)	13 (09)	11 (08)	26 (21)	35 (25)	11 (09)	ns	23 (17)	73
Reliability	68	61	79	91	68	88	92	73	

Note. Correlations are corrected for attenuation, with uncorrected correlations in parentheses. Leadi decimals omitted. *N*s range from 423 to 1,353. All correlations are significant at the .01 level.

[a]Notation for worksamples: GC (Gestalt Completion); SP (Snowy Pictures); CW (Concealed Words); HP (Hidden Patterns); HF (Hidden Figures); DR (Drawing); MZ (Maze Tracing Speed Test); TC (Tactics).

Of the eight experimental tests, Hidden Patterns, Hidden Figures, and Drawing exhibited the greatest number of moderate and high correlations with Foundation worksamples. In addition, the similar pattern of correlations for Tactics and the flexibility-of-closure tests with the JOCRF cognitive worksamples suggests that Tactics can reasonably be classified as a measure of flexibility of closure and not spatial scanning. (From this point on, Tactics will be referred to as a flexibility-of-closure test rather than a spatial scanning measure.) All four of these perceptual abilities tests correlated at a moderate to high level with Paper Folding, Wiggly Block, Memory for Design, Number Series, and Analytical Reasoning. The correlations with Paper Folding and Wiggly Block indicate that flexibility of closure has a substantial relationship with structural visualization. The correlations with Memory for Design, Number Series, and Analytical Reasoning may also be due to structural visualization, because Paper Folding and Wiggly Block correlate at least moderately with Memory for Design (rs = .69 and .59, respectively), Number Series (rs = .52 and .41, respectively), and Analytical Reasoning (rs = .60 and .54, respectively). In addition, this common relationship with structural visualization may be responsible for the moderate correlations of Hidden Figures and Drawing with Mathematics Vocabulary (which correlates .52 with Paper Folding and .35 with Wiggly Block). Moderate correlations were also found for Hidden Patterns with Inductive Reasoning and Number Checking, both of which, like Hidden Patterns, are highly speeded worksamples. The only other moderate relationships between the flexibility-of-closure measures and the cognitive aptitude tests were found for Hidden Figures with English Vocabulary and Reading Efficiency.

All three of the speed-of-closure markers correlated moderately with Wiggly Block, Paper Folding, and Memory for Design. In addition, Gestalt Completion correlated with Inductive and Analytical Reasoning, Snowy Pictures with Inductive Reasoning and Observation, and Concealed Words with Analytical Reasoning and Number Series.

Maze Tracing did not correlate above .20 with any of the cognitive measures, including structural visualization. This was unexpected in light of previous Foundation research findings that Map Planning, another ETS spatial-scanning measure, correlated .49 and .45 with Paper Folding and Wiggly Block, respectively (Technical Report 1981-2).

Initial principal components analysis with pairwise deletion for the perceptual abilities tests and the cognitive worksamples extracted six factors with eigenvalues greater than or equal to unity, namely, 6.09, 2.36, 1.66, 1.34, 1.08, and 1.00. The total variance attributable to each of the first four factors was 27.7%, 10.7%, 7.6%, and 6.1%, respectively. None of the remaining factors accounted for more than 5% of the variance.

Application of the scree test would have reduced the number of factors from six to four, but the present study was more interested in illuminating the relationships among the worksamples than specifying the most-parsimonious factor structure. Accordingly, 4-, 5-, and 6-factor solutions were rotated to the varimax criterion. Because Maze Tracing did not correlate even moderately with any of the worksamples in either the experimental or standard battery, this test was eliminated from the factor analyses. The 5-factor structure seemed more interpretable than the 4-factor solution; specifically, the latter solution appeared to force multiple factors together. At the same time, no additional perceptual abilities factors were manifested in the 6-factor solution. For these reasons, the rotated factor matrix for five factors is reported in Table 18.

Factor 1 involves the flexibility-of-closure measures, together with the spatially related JOCRF worksamples. Factors 2 and 3 can be interpreted as memory and verbal-facility factors, respectively. The highest loadings for Factor 4 were related to measures of idea production, while the highest loadings for Factor 5 encompassed speed-related aptitudes, particularly speed of closure.

Based on the results of this factor analysis together with the correlation patterns, it is evident that the flexibility-of-closure and speed-of-closure tests form factors independent of each other. Furthermore, the speed-of-closure markers remain together and are more or less distinct from the other cognitive worksamples. Similarly, the flexibility-of-closure tests remain together when factored with the cognitive worksamples, but it is important to note that other cognitive worksamples also loaded on the same factor as the flexibility-of-closure measures. Specifically, there is evidence of a relatively strong relationship between flexibility of closure and spatial aptitude measures. This suggests that Hidden Patterns, Hidden Figures, Drawing, and Tactics do not measure flexibility of closure exclusively. Moreover, because Factor 1 has substantial loadings on tests other than structural visualization, the factor appears not to be measuring just a structural component. Instead, what is most likely being measured is a spatial component common to both flexibility of closure and structural visualization.

Three additional comments with regard to the flexibility-of-closure worksamples are in order. First, despite the relatively strong relationships of Hidden Patterns and Hidden Figures with structural visualization, the sex effects of the former were quite different from the latter; that is, no sex differences were found for Hidden Patterns and Hidden Figures, whereas males significantly outperformed females on Wiggly Block and Paper Folding, $\underline{F}(1;1,267) = 69.90$, $\underline{p} < .01$, $\underline{w}^2 = .05$, and $\underline{F}(1;1,267) = 42.87$, $\underline{p} < .01$, $\underline{w}^2 = .03$, respectively. This absence of a sex difference is apparently due to the nonstructural component in flexibility of closure. Second, even

Table 18

Factor Analysis of the Perceptual Abilities Tests and Cognitive Worksamples in JOCRF Battery[a]

Factor loadings for rotated[b] five-factor solution

Worksample	F1	F2	F3	F4	F5
Wiggly Block	.78	.14	-.06	.03	.14
Paper Folding	.74	.23	.18	-.11	.19
Memory for Design	.49	.56	.09	.00	.26
Silograms	-.06	.67	.38	.07	.09
Number Memory	.13	.68	.20	-.04	.16
Observation	.14	.73	-.15	.18	.03
English Vocabulary	.06	-.02	.77	.32	-.05
Reading Efficiency	.05	.14	.52	.55	.06
Number Series	.39	.33	.50	.13	.15
Math Vocabulary	.33	.26	.71	.04	.05
Ideaphoria	-.02	.07	.16	.69	-.10
Foresight	.08	-.06	.11	.65	-.07
Graphoria	-.16	.20	.19	.49	.36
Inductive Reasoning	.14	.30	-.22	.54	.40
Analytical Reasoning	.44	.38	.24	.28	.10
Gestalt Completion	.24	.03	.00	-.07	.69
Snowy Pictures	.08	.20	-.08	.06	.70
Concealed Words	.23	.07	.21	-.08	.72
Hidden Patterns	.42	.11	.11	.34	.45
Hidden Figures	.48	.03	.36	.14	.25
Drawing	.65	-.03	.23	-.12	.15
Tactics	.72	.04	.02	.13	.05

Note. Pairwise deletion of missing cases. *N*s range from 423 to 1,353.

[a] Initial extraction using principal components analysis.
[b] Varimax rotation.

though Drawing and Tactics are more strongly related to structural visualization than are Hidden Patterns and Hidden Figures, the latter two tests do not form their own factor. Third, Hidden Patterns and Hidden Figures have lower correlations with Wiggly Block and Paper Folding and smaller loadings on Factor 1 than does Memory for Design, a worksample that is regarded as a measure of an aptitude distinct from structural visualization.

Perceptual Abilities and the Noncognitive Worksamples

Table 19 presents the correlations of the eight tests in the perceptual abilities battery with the noncognitive worksamples in the JOCRF standard battery. None of the experimental measures displayed moderate or high correlations with the noncognitive worksamples. The strongest relationships found were for Hidden Patterns with the music tests and Finger Dexterity and for Hidden Figures with Tonal Memory and Rhythm Memory.

Predicting Perceptual Abilities Scores from Scores on the JOCRF Standard Worksamples

In making a decision regarding which, if any, of the perceptual abilities tests to add to the standard battery, one question that should be addressed is: Do any of the experimental tests represent needless duplication with standard worksamples or do they measure constructs unique to the perceptual abilities battery? Based on the relatively low correlations between the Maze Tracing Speed Test and the Foundation worksamples, it appears that the construct measured by Maze Tracing is unique; that is, scores for Maze Tracing cannot be estimated from scores on worksamples in the standard battery. Most of the correlations of the speed-of-closure markers with the Foundation worksamples are also relatively low; still, there are several moderate correlations between these perceptual abilities tests and the standard worksamples, which suggests the existence of some overlap between speed of closure and the constructs measured by the JOCRF battery. The moderate correlations between the flexibility-of-closure measures and the spatially related worksamples indicate that the construct measured by the flexibility-of-closure tests is partly duplicated by worksamples in the standard battery.

To determine the amount of overlap between the perceptual abilities measures and the JOCRF worksamples, each perceptual abilities test was regressed on all the worksamples in the standard battery. The Foundation worksamples were entered into the analyses hierarchically, with the cognitive worksamples entered in the first step and the noncognitive tests in the second step. Table 20 presents the results of these regression analyses.

As expected, the greatest overlap between the experimental and standard worksamples was found for the flexibility-of-closure

Table 19

Correlations Between Perceptual Abilities Tests and Noncognitive Worksamples in JOCRF Battery

Worksample[a]	GC	SP	CW	HP	HF	DR	MZ	TC	Rel.
Color Perception	—	—	—	—	—	—	—	—	—
Personality	ns	ns	ns	ns	ns	ns	ns	ns	89
Tonal Memory	16 (13)	17 (13)	20 (17)	24 (22)	28 (22)	14 (13)	ns	ns	92
Pitch Discrim.	20 (15)	17 (12)	21 (17)	28 (24)	24 (18)	21 (18)	ns	24 (18)	80
Rhythm Memory	16 (11)	16 (11)	26 (20)	26 (21)	30 (21)	24 (19)	ns	22 (16)	73
Finger Dexterity	ns	22 (16)	ns	28 (25)	24 (18)	ns	ns	ns	86
Tweezer Dexterity	ns	15 (11)	ns	20 (18)	13 (10)	ns	ns	ns	93
Writing Speed	—	—	—	— (16)	— (09)	—	—	—	—
Reliability	68	61	79	91	68	88	92	73	

Note. Correlations are corrected for attenuation, with uncorrected correlations in parentheses. Leading decimals omitted. Ns range from 423 to 1,353. All correlations are significant at the .01 level.

[a]Notation for worksamples: GC (Gestalt Completion); SP (Snowy Pictures); CW (Concealed Words); HP (Hidden Patterns); HF (Hidden Figures); DR (Drawing); MZ (Maze Tracing Speed Test); TC (Tactics).

Table 20

Regression Analyses Predicting Perceptual Abilities Worksamples Using Worksamples from JOCRF Standard Battery

Worksample	Relia.	Step[a]	Multiple R[b]	R-squared[c]	Percentage of unique, reliable variance[c]
Gestalt Completion	.68	Step 1	.45 (.42)	.18	.50
		Step 2	.46 (.43)		
Snowy Pictures	.61	Step 1	.42 (.40)	.16	.45
		Step 2	.43 (.40)		
Concealed Words	.79	Step 1	.47 (.45)	.20	.59
		Step 2	.48 (.46)		
Hidden Patterns	.91	Step 1	.58 (.57)	.32	.59
		Step 2	.60 (.58)		
Hidden Figures	.68[d]	Step 1	.55 (.54)	.29	.39
		Step 2	.56 (.54)		
Drawing	.88	Step 1	.56 (.55)	.30	.58
		Step 2	.57 (.55)		
Maze Tracing	.92	Step 1	.30 (.26)	.07	.85
		Step 2	.31 (.26)		
Tactics	.73	Step 1	.59 (.57)	.32	.41
		Step 2	.62 (.57)		

[a] In Step 1 the cognitive worksamples were entered into the analyses; in Step 2 the noncognitive worksamples were entered.

[b] Number in parenthesis is multiple correlation adjusted for shrinkage.

[c] R-squared and unique reliable variance were not calculated for Step 2, because the noncognitive worksamples explained relatively little variance in the perceptual abilities tests, after the variance explained by the cognitive worksamples was removed. R-squared is based on the multiple R adjusted for shrinkage.

[d] Estimated reliability for Part 1 only of Hidden Figures was computed from reported reliabilites for the whole test of .80 for females and .82 for males (Ekstrom et al., 1976), using the Spearman-Brown formula for reliability of a test of length .5.

tests: the multiple Rs ranged from .54 to .58, which indicates that scores on Hidden Patterns, Hidden Figures, Drawing, and Tactics were moderately well predicted from the standard battery. Nevertheless, only about 29 to 34% of the variance in these tests is explained by all the cognitive and noncognitive worksamples in the standard battery (see the R-squared values in Table 20). This means that the construct of flexibility of closure cannot be measured without administering at least one of the flexibility-of-closure tests in the perceptual abilities battery.

For the speed-of-closure markers, there is some, but not much, overlap with the Foundation worksamples. Approximately 16 to 21% of the variance in Gestalt Completion, Snowy Pictures, and Concealed Words is explained by all the JOCRF worksamples combined, which confirms that speed of closure is relatively independent of the aptitudes measured by the Foundation.

Not surprisingly, very little of the variance in the Maze Tracing Speed Test is attributable to the worksamples in the standard battery. In other words, the construct measured by Maze Tracing is unique to the perceptual abilities battery.

Also as can be seen in Table 20, little variance in the perceptual abilities measures is explained by the noncognitive tests in the JOCRF standard battery, after the variance explained by the cognitive worksamples is removed. This was not surprising in light of the relatively low correlations between the perceptual abilities and the noncognitive worksamples.

The last column in Table 20 presents the unique, reliable variance in each of the perceptual abilities measures--that is, the variance that cannot be explained by the Foundation worksamples when the reliabilities of the experimental tests are taken into account. The unique, reliable variance in the flexibility-of-closure tests ranges from approximately 39 to 59%. Both Tactics and Hidden Figures have a moderate amount of redundancy with the cognitive worksamples and somewhat low reliability; therefore, they contain less unique, reliable variance than the other two measures of flexibility of closure, Hidden Patterns and Drawing. For the speed-of-closure markers, the reliable variances not explained by the JOCRF standard battery are about 45%, 50%, and 59% for Snowy Pictures, Gestalt Completion, and Concealed Words, respectively. The percent of unique, reliable variance for Maze Tracing is very high, 85%.

In summary, there is substantial overlap between Hidden Patterns, Hidden Figures, Drawing, and Tactics and the cognitive worksamples in the standard battery, but these measures of flexibility of closure also have a moderate amount of variance unique to themselves. Because of the lower reliabilities of Hidden Figures and Tactics, Hidden Patterns and Drawing appear to be the best of the four flexibility-of-closure tests. The speed-of-closure markers are relatively independent of the JOCRF worksamples. Compared to Gestalt Completion and Snowy Pictures, Concealed Words, with its higher reliability, has a greater percentage of reliable variance that cannot be explained by the standard batt . The construct measured by the Maze Tracing Speed Test is unique.

Perceptual Abilities and Laterality, Education, and College Major

Perceptual Abilities and Laterality

As previously mentioned in the Method section, a multivariate analysis of variance was performed with all the experimental tests except Tactics as dependent variables and sex, eyedness, and handedness as independent variables. The multivariate test for handedness main effects was not significant, nor were any of the tests for interaction effects (i.e., eyedness by sex, handedness by sex, eyedness by handedness, eyedness by handedness by sex). As expected, the sex main effect was significant, Wilks' lambda = .94, $p < .001$. The multivariate test for eyedness main effects was also significant, Wilks' lambda = .98, $p < .05$. The univariate F-tests for eyedness revealed a significant effect for Drawing ($p = .014$), with the effect for Hidden Figures approaching significance ($p = .079$). This is consistent with the findings from a recent Foundation laterality study by Zinbarg (Technical Report 1987-2), in which eyedness effects were found principally on tests thought to be related to spacial ability.

For the 10-trial Tactics worksample, an analysis of variance of test scores by sex, eyedness, and handedness showed a significant sex effect, $F(1,413) = 45.76$, $p < .001$. Neither the eyedness nor handedness effect was significant, and no significant interaction was found.

For each of the perceptual abilities worksamples, the means for the eyedness groups for males and females separately are presented in Table 21. On the Drawing test, the variable-eyed group performed significantly better than the left- and right-eyed groups. A similar trend was observed for the other experimental worksamples, even though their univariate F-tests for eyedness were not significant. For the flexibility-of-closure measures and the Maze Tracing Speed Test, the variable-eyed group tended to perform better than the other two groups among both males and females. In addition, for the flexibility-of-closure measures, there was a tendency for left-eyed females to score lower than either variable- or right-eyed females. For the speed-of-closure marker tests, left-eyed females tended to score lower than either variable-eyed or right-eyed females; among males, the trend was for the variable-eyed group to outperform the left- and right-eyed groups. These results are congruent with Zinbarg's (Technical Report 1987-2) finding that variable-eyed males significantly

Table 21

Mean Performance on Perceptual Abilities Tests for Female and Male Eyedness Groups

Worksample	Females: All	Females Eyedness group: Left	Females Eyedness group: Variable	Females Eyedness group: Right	Males: All	Males Eyedness group: Left	Males Eyedness group: Variable	Males Eyedness group: Right	Overall SD
Gestalt Completion	13.94	13.44	13.64	14.33	14.69	14.80	14.76	14.60	2.86
Snowy Pictures	16.40	16.07	16.51	16.48	16.41	16.08	16.77	16.27	3.37
Concealed Words	25.22	24.39	25.37	25.52	26.70	25.73	28.02	26.12	7.52
Hidden Patterns	195.69	191.70	199.61	195.28	196.72	191.74	203.49	193.76	41.61
Hidden Figures	6.68	6.28	6.93	6.72	6.79	6.70	7.43	6.34	4.49
Drawing	23.34	21.28	24.97	23.36	31.58	29.41	34.38	30.37*	15.24
Tactics	49.21	50.38	52.10	46.90	63.63	64.16	65.81	61.56	22.17
Maze Tracing	29.31	29.08	30.54	28.74	32.41	32.05	32.93	32.17	8.76
n	595	131	164	300	635	120	218	297	

*The effect for eyedness was significant at the .05 level (occurred only for the Drawing test).

outperform the other two groups on a number of cognitive ability tests, including spatially related aptitude measures, with a similar trend displayed among females.

Perceptual Abilities and Educational Level

The multivariate analysis of variance using educational level at time of testing as the independent variable yielded a significant main effect, Wilks' lambda = .79, $p < .001$. The univariate F-tests were significant at the .05 level for Concealed Words, Hidden Patterns, Hidden Figures, and Maze Tracing. The main effect was not significant for the analysis of variance of Tactics scores by education.

Because the majority of Foundation examinees are in high school or college at the time of their testing and are still in the process of completing their schooling, a second multivariate analysis of variance using only adults aged 23 or older was performed. Again, a significant main effect was found, Wilks' lambda = .82, $p < .05$, with the univariate F-tests significant at the .01 level for Concealed Words and Hidden Figures and approaching significance at the .05 level for Hidden Patterns ($p = .054$). In the analysis of variance of Tactics scores by educational level for examinees older than 22, no significant main effect was found. The pattern that emerges is that adults with at least some education beyond high school perform better on Concealed Words, Hidden Patterns, and Hidden Figures than those individuals with only a high school education.

Perceptual Abilities and College Major

Still to be addressed is whether the perceptual abilities have criterion-related validity with regard to non-test data such as biographical information. A variable that was available for exploring this question was examinees' choice of college major, as described earlier in the Method section. Of interest was whether individuals who excel in speed of closure, flexibility of closure, and spatial scanning would tend to select and persist in certain major fields of study, while those who score low in the aptitude would tend to avoid those college majors.

To assess the relationship of college major to the perceptual abilities, a multivariate analysis of variance was performed, with the experimental tests (excluding Tactics) as the dependent variables and major field of study as the independent variable. The multivariate test for main effects was significant, Wilks' lambda = .76, $p < .001$. Two more multivariate analyses of variance were then performed: one with only the speed-of-closure markers as the dependent variables and a second with the flexibility-of-closure measures (excluding Tactics) as the dependent variables. No significant main effect was found for the former analysis, indicating that scores on Gestalt Completion, Snowy Pictures, and Concealed Words are not significantly different across college-major categories. With regard to the latter analysis, the multivariate test for main effects was significant, Wilks' lambda = .82, $p < .001$. The univariate analyses of variance revealed significant effects for all three tests (i.e., Hidden Patterns, Hidden Figures, and Drawing). For the 10-trial version of Tactics, an analysis of variance of test scores by college major also showed a significant main effect, $F(6,129) = 5.05$, $p < .001$. In an analysis of variance of scores on Maze Tracing by college major, the main effect was not significant. To review, the multivariate and univariate analyses of variance indicated significant relationships with college major for all four of the flexibility-of-closure measures and not for the speed-of-closure markers or Maze Tracing.

The mean scores for the college-major groups are presented in Table 22. Two striking results emerged from these analyses. The first is that engineering majors scored considerably higher than the other college-major groups on the flexibility-of-closure measures, with the exception of Hidden Patterns. The second is the similarity in the ordering of the groups across the perceptual abilities; that is, engineers along with biological and physical science majors score highest on these worksamples, then business and social science majors, followed by humanities and education majors. It thus appears that individuals who choose majors in the fields of engineering, physical science, and biological science tend to score higher on the perceptual abilities worksamples, particularly the flexibility-of-closure tests, than those who choose other majors.

Because of the overlap found between the flexibility-of-closure measures and the Foundation's spatially related tests, analyses of variance for the JOCRF structural visualization worksamples by college major were performed. The main effect was significant for both Wiggly Block and Paper Folding, $F(6,439) = 7.53$, $p < .001$, and $F(6,451) = 11.11$, $p < .001$, respectively. Mean scores on these two Foundation worksamples for the seven college-major groups are also presented in Table 22.

It appears that flexibility of closure contributes to the same majors as structural visualization, which suggests that there may be a mediating effect operating among the structural-visualization and flexibility-of-closure measures. Because of the relationship between flexibility of closure and structural visualization and the relationship between structural visualization and college majors (which follows the same pattern as the relationship between flexibility of closure and major field of study), structural visualization could account for part of the relationship between flexibility of closure and college majors; but, statistically, structural visualization can only account for a fraction of that relationship. In other words, at least part of the contribution of flexibility of closure is distinct from that of structural visualization.

Table 22

Mean Performance on Perceptual Abilities Tests by College Major

Worksample	College-major category[a] 1	2	3	4	5	6	7	Over-all SD
Gestalt Completion	14.78	15.00	15.47	14.09	14.29	13.93	13.17	2.98
Snowy Pictures	16.53	16.95	17.60	16.51	16.36	16.12	15.14	3.50
Concealed Words	27.54	27.06	28.93	26.60	26.73	25.50	24.61	7.62
Hidden Patterns	215.78	229.05	219.90	201.09	200.34	194.24	191.18*	43.21
Hidden Figures	10.74	8.81	8.81	7.36	6.91	6.33	6.24*	4.39
Drawing	41.65	32.35	29.72	26.[illegible]	27.29	24.18	24.03*	15.05
Tactics	83.62	61.65	55.96	59.46	51.55	49.27	49.25*	21.23
Maze Tracing	30.96	32.67	32.38	30.03	28.43	24.98	27.62	15.55
Wiggly Block	325.91	304.14	287.00	265.92	254.12	232.91	217.81*	99.24
Paper Folding	33.31	28.70	25.13	22.11	20.19	17.66	15.78*	13.43
n	46	21	32	151	80	97	45	

[a]Notation for college-major categories: 1 (Engineering); 2 (Physical sciences); 3 (Biological sciences); 4 (Business); 5 (Social sciences); 6 (Humanities); 7 (Education).

*The effects from the univariate F-tests were significant at the .001 level.

GENERAL SUMMARY AND CONCLUSIONS

The purpose of this study was to evaluate the psychometric properties of the individual worksamples in the perceptual abilities battery and to assess the relationships among the eight experimental worksamples, as well as the relationships of the perceptual abilities to the aptitudes measured by the JOCRF standard battery. The eight experimental worksamples considered jointly were found to measure three distinct factors, with one represented by Maze Tracing (tentatively labeled "spatial scanning"), a second by Gestalt Completion, Snowy Pictures, and Concealed Words (labeled "speed of closure"), and a third by Hidden Patterns, Hidden Figures, Drawing, and Tactics (labeled "flexibility of closure").

It was also demonstrated that: (a) the Maze Tracing Speed Test was not related to any of the JOCRF worksamples (i.e., this measure showed good discriminant validity), (b) the speed-of-closure markers were relatively independent of the JOCRF worksamples, and (c) the flexibility-of-closure measures had both substantial overlap with the spatially related worksamples in the standard battery and a moderate amount of variance unexplained by the Foundation worksamples.

The Maze Tracing Speed Test is a highly reliable worksample that takes but a few minutes to administer. Maze Tracing not only does not relate to any of the other tests in the perceptual abilities battery, including Tactics, but also does not correlate even moderately with any of the cognitive or noncognitive worksamples in the standard battery, including structural visualization. Whether Maze Tracing represents a more general perceptual ability, such as spatial scanning, however, was not possible to determine in this study and remains to be investigated.

Among the speed-of-closure markers, both Gestalt Completion and Snowy Pictures had reliabilities that were lower than desirable. As discussed in an earlier section of this report, one way to increase a worksample's reliability to an acceptable .80 is to lengthen it by adding items. Gestalt Completion, to be reliable, would need to be almost doubled in length, which would extend total test time to approximately eight minutes, whereas a reliable version of Snowy Pictures would require over 20 minutes to administer. For this reason a lengthened version of Gestalt Completion would be recommended over a longer version of Snowy Pictures. It is important to note, however, that the process of adding items to a test to increase its reliability entails not only designing additional items but also performing another set of item analyses to verify their effectiveness.

Concealed Words, on the other hand, had a split-half reliability near the Foundation's recommended minimum standard of

.80 and required eight minutes to administer (not including instructions). Compared to Gestalt Completion and Snowy Pictures, Concealed Words also had a greater percentage of reliable variance that could not be explained by the JOCRF standard battery. A potential drawback to the use of Concealed Words rather than Gestalt C[illegible]n or Snowy Pictures in the JOCRF standard battery is [illegible] Concealed Words Test involves words rather than pictures and, [illegible]efore, might appear (e.g., to examinees) to be more related to verbal than perceptual abilities. Still, on the basis of psychometric properties, such as reliability, as well as practical considerations, such as ease and speed of administration, Concealed Words is superior to the other two markers of speed of closure.

A factor measuring flexibility of closure exclusively did not emerge in this study. Instead, both the flexibility-of-closure and structural visualization worksamples loaded on the same factor. It was determined, however, that flexibility of closure is measuring a construct distinct from that of structural visualization. This was based on: (a) examination of the correlation pattern, which suggested that a spatial component common to both flexibility of closure and structural visualization is most likely responsible for the moderate to high correlations between the flexibility-of-closure and spatially related worksamples; and (b) the fact that there was a moderate amount of variance in flexibility of closure that could not be explained by the Foundation worksamples, which indicated that the construct of flexibility of closure could not be measured sufficiently by the JOCRF battery alone; that is, without administering at least one of the flexibility-of-closure tests in the perceptual abilities battery.

With regard to the individual flexibility-of-closure measures, Hidden Patterns and Drawing possessed adequate reliabilities, while Hidden Figures and Tactics had lower-than-desirable reliabilities. This means that significant revision of Hidden Figures or Tactics would be necessary before they were used further by the Foundation. In addition, even though the R-squared values are similar for all four of the flexibility-of-closure worksamples, the reliable variance in the flexibility-of-closure tests that could not be explained by all the Foundation worksamples combined was greater for Hidden Patterns and Drawing than for Hidden Figures and Tactics, indicating that the former tests make more of a unique contribution than the latter. Although all four flexibility-of-closure measures correlated moderately with a number of cognitive worksamples in the standard battery, particularly the spatially related measures, Hidden Patterns and Hidden Figures were less related than Drawing and Tactics to structural visualization. On the basis of test reliability, unique reliable variance, and relationships with the worksamples of the standard battery, Hidden Patterns had the best discriminant validity of the flexibility-of-closure measures. It was therefore concluded that Hidden Patterns measures the construct of flexibility of closure somewhat better (or more purely) than the other experimental perceptual abilities tests.

Conclusions

The results of this study indicated that, in general, the perceptual abilities factors of flexibility of closure, speed of closure, and spatial scanning are not measured by the worksamples in the JOCRF standard battery. In other words, the incorporation of one or more measures of these perceptual abilities into the Foundation battery would not represent needless duplication with the standard aptitude worksamples. In particular, the analyses suggested that: (a) the construct measured by Maze Tracing represents an independent cognitive factor, perhaps spatial scanning; and (b) Concealed Words and Hidden Patterns display better discriminant validity than the other experimental speed-of-closure and flexibility-of-closure worksamples.

Of the tests in the experimental battery, then, the Concealed Words Test, Hidden Patterns Test, and Maze Tracing Speed Test can be considered the most suitable perceptual abilities measures. If further research indicates them to be predictive of occupational preference and job satisfaction, these measures may make useful contributions to the JOCRF standard battery.

REFERENCES

Adcock, C. J., & Martin, W. A. (1971). Flexibility and creativity. Journal of General Psychology, 85, 71-76.

Anastasi, A. (1976). Psychological testing (4th ed.). New York: Macmillan.

Barrett, H. C., & Barker, H. R. (1973). Cognitive pattern perception and musical performance. Perceptual and Motor Skills, 36, 1187-1193.

Baten, L. (1981). Test comprehension. The parameters of difficulty in narrative and expository prose texts: A redefinition of readability. Research report, University of Illinois.

Bechtoldt, H. P. (1947). Factorial investigation of the perceptual speed factor. American Psychologist, 2, 304-305.

Bejar, I. I., & Yocum, P. (1986). Analysis and generation of hidden figure items: A cognitive approach to psychometric modeling. (RR-86-19). Princeton, NJ: Educational Testing Service.

Botzum, W. A. (1951). A factorial study of the reasoning and closure factors. Psychometrika, 16, 361-386.

Bouma, A., & Ippel, M. J. (1983). Individual differences in mode of processing in visual asymmetry tasks. Cortex, 19, 51-68.

Bunderson, C. V. (1967). Transfer of mental abilities at different stages of practice in the solution of concept problems. (ETS RB 67-20). Princeton, NJ: Educational Testing Service.

Cohen, J., & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Crosson, C. W. (1984). Age and field independence among women. Experimental Aging Research, 10, 165-170.

Cunningham, W. R. (1978). On the generality of ability factor structure across the adult life-span. Research report, University of Florida.

Cunningham, W. R. (1981). Ability factor structure differences in adulthood and old age. Multivariate Behavioral Research, 16, 3-22.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. Multivariate Behavioral Research Monographs, 79(2), 1-84.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). Manual for Kit of Factor-Referenced Cognitive Tests. Princeton, NJ: Educational Testing Service.

French, J. W. Ekstrom, R. B., & Price, L. A. (1963). Kit of Reference Tests for Cognitive Factors. Princeton, NJ: Educational Testing Service.

Fry, K. L., & Thompson, D. S. (1977). Comparison of selected measures of field dependence. Perceptual and Motor Skills, 45, 861-862.

Glass, G. G. (1967). Rate of reading: A correlation and treatment study. Journal of Reading, 11, 168-178.

Guilford, J. P., & Lacey, J. I. (Eds.). (1947). Printed classification tests. Washington, DC: U.S. Government Printing Office.

Hill, N. R. (1970, November). Eight-line Drawing, Worksample 494. Paper presented at the research meeting of the Human Engineering Laboratory-Johnson O'Connor Research Foundation, New York.

Irons, D. M. (1982). Predicting programming performance in novice programmers by measures of cognitive abilities. Unpublished doctoral dissertation, Texas Christian University.

Johnson O'Connor Research Foundation. (1968). Administration of experimental worksamples for the 5-appointment battery. Unpublished document.

Kaplan, K. J. (1986). Perceptual abilities: A review of the literature. Unpublished manuscript.

Kastner, M. (1978). A contribution to the clarification of the cognitive components of intelligence. Diagnostica, 24, 124-136.

Lemke, E., Klausmeier, H. J., & Harris, C. W. (1967). Relationship of selected cognitive abilities to concept attainment and information processing. Journal of Educational Psychology, 58, 27-35.

Linn, M. C., & Kyllonen, P. (1981). The field dependence-independence construct: Some, one, or none. Journal of Educational Psychology, 73, 261-273.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. Child Development, 56, 1479-1498.

Maccoby, E. E., & Jacklin, C. N. (1974). The psychology of sex differences. Stanford, CA: Stanford University Press.

Messick, S., & French, J. W. (1975). Dimensions of cognitive closure. Multivariate Behavioral Research, 10, 3-16.

Minutes of Meeting of the Research Department and Executive Committee. (1971, November). Unpublished document.

Minutes of Meeting of the Trustees. (1964, June). Unpublished document.

Minutes of Meeting of the Trustees. (1966, May). Unpublished document.

Mos, L., Wardell, D., & Royce, J. R. (1974). A factor analysis of some measures of cognitive style. Multivariate Behavioral Research, 9, 47-57.

Norusis, M. J. (1988a). SPSS/PC+ V2.0 base manual. Chicago: SPSS.

Norusis, M. J. (1988b). SPSS/PC+ Advanced Statistics V2.0. Chicago: SPSS.

Nunnally, J. C. (1967). Psychometric theory. New York: McGraw-Hill.

Ohnmacht, F. W., & Fleming, J. T. (1972). Perceptual closure and cloze performance: A replication with older subjects. Journal of General Psychology, 87, 225-229.

Ohnmacht, F. W., Weaver, W. W., & Kohler, E. T. (1970). Cloze and closure: A factorial study. Journal of Psychology, 74, 205-217.

Parasuraman, R. (1976). Consistency of individual differences in human vigilance performance: An abilities classification analysis. Journal of Applied Psychology, 61, 486-492.

Pemberton, C. L. (1952). The closure factors related to other cognitive processes. Psychometrika, 17, 267-288.

Poff, R., & Holder, B. (November, 1974). Drawing research, Worksample 494C, leading to Form D: Determination of order of difficulty. Paper presented at the reasearch meeting of the Human Engineering Laboratory-Johnson O'Connor Research Foundation, New York.

Richmond, P. G. (1980). A limited sex difference in spatial test scores with a preadolescent sample. Child Development, 51, 601-602.

Roff, M. A. (1953). A factorial study of tests in the perceptual area. Psychometric Monographs, No. 8. Chicago: University of Chicago Press.

Royce, J. R. (1973). The conceptual framework for a multi-factor theory of individuality. In J. R. Royce (Ed.), Multivariate analysis and psychological theory (pp. 305-407). New York: Academic Press.

Sawa, H. (1969). Analytic thinking and synthetic thinking: A research through perception. Psychologia, 12, 3-16.

Shute, V. J., Pellegrino, J. W., Hubert, L., & Reynolds, R. W. (1983). The relationship between androgen levels and human spatial abilities. Bulletin of the Psychonomic Society, 21, 465-468.

Spearman, C. E. (1927). The abilities of man. New York: Macmillan.

Tarter, R. E., Sheldon, J., & Sugarman, A. A. (1975). Correlates of perceptual field orientation in alcoholics. Journal of Clinical Psychology, 31, 364-366.

Technical Report 168. (1945). A new experiment in the measurement of some reasoning other than inductive or analytical. L. I. Kimball. Boston: Johnson O'Connor Research Foundation.

Technical Report 1981-2. Construct validity of the Wiggly Block and Paper Folding tests: A factor-referenced multivariate study. R. S. Trafton. Boston: Johnson O'Connor Research Foundation.

Technical Report 1982-3. Male-female differences in cognitive abilities in various age groups. R. S. Trafton & W. T. Garrison. Boston: Johnson O'Connor Research Foundation.

Technical Report 1987-1. The numerical facility project. J. Tal. Chicago: Johnson O'Connor Research Foundation.

Technical Report 1987-2. Sidedness, health disorders, and aptitudes. R. Zinbarg. Chicago: Johnson O'Connor Research Foundation.

Test Information Bulletin 1980-7. Reliabilities and standard errors of laboratory worksamples. M. Daniel. Chicago: Johnson O'Connor Research Foundation.

Thurstone, L. L. (1944). A factorial study of perception. Psychometric Monographs, No. 4. Chicago: University of Chicago Press.

Thurstone, L. L. (1947). Multiple factor analysis. Chicago: University of Chicago Press.

Tyler, L. E. (1956). The psychology of human differences (2nd ed.). New York: Appleton-Century-Crofts.

Wallace, J. P., & Mandell, M. (1972, May). Reliability study: Video tape 8-line Drawing. Paper presented at the research meeting of the Human Engineering Laboratory-Johnson O'Connor Research Foundation, New York.

Wiggly Block Worksample 3,4,5,6,7 Manual. (1978). M. Daniel. Chicago: Johnson O'Connor Research Foundation.

Witkin, H. A. (1949). The nature and importance of individual differences in perception. Journal of Personality, 18, 145-170.

Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R., & Karp, S. A. (1962). Psychological differentiation. New York: Wiley.

Witkin, H. A., & Goodenough, D. R. (1981). Cognitive styles: Essence and origins. New York: International Universities Press.

Wright, B. D., Congdon, R. T., & Rossner, M. (1987). MSCALE: A Rasch program for rating scale analysis. Chicago: University of Chicago, Department of Education, Statistical Laboratory.

JOHNSON O'CONNOR RESEARCH FOUNDATION

HUMAN ENGINEERING LABORATORY